



92

500.43106X00

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): Hiroki KANAI
Serial No.: 10/654,996
Filed: September 5, 2003
Title: STORAGE DEVICE CONTROL APPARATUS AND CONTROL
METHOD FOR THE STORAGE DEVICE CONTROL APPARATUS

LETTER CLAIMING RIGHT OF PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

September 25, 2003

Sir:


Under the provisions of 35 USC 119 and 37 CFR 1.55, the applicant(s) hereby
claim(s) the right of priority based on:

Japanese Patent Application No. 2003-111405
Filed: April 16, 2003

A certified copy of said Japanese Patent Application is attached.

Respectfully submitted,

ANTONELLI, TERRY, STOUT & KRAUS, LLP



Carl I. Brundidge
Registration No. 29,621

CIB/rp
Attachment

日本国特許庁 W1160-01EW
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年 4月16日
Date of Application:

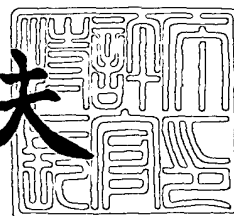
出願番号 特願2003-111405
Application Number:
[ST. 10/C]: [JP 2003-111405]

出願人 株式会社日立製作所
Applicant(s):

2003年 8月26日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



出証番号 出証特2003-3069635

【書類名】 特許願

【整理番号】 HI030082

【提出日】 平成15年 4月16日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 3/06

【発明者】

【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社日立製作所 R A I D システム事業部内

【氏名】 金井 宏樹

【特許出願人】

【識別番号】 000005108

【氏名又は名称】 株式会社日立製作所

【代理人】

【識別番号】 110000176

【氏名又は名称】 一色国際特許業務法人

【代表者】 一色 健輔

【手数料の表示】

【予納台帳番号】 211868

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 記憶デバイス制御装置、及び記憶デバイス制御装置の制御方法

【特許請求の範囲】

【請求項 1】 データ入出力要求を受信するためのホストインタフェース制御部が形成されたチャネル制御ユニットと、

前記データ入出力要求に応じて、データを記憶するための記憶ボリュームに対する前記データの入出力制御を行うためのディスクインタフェース制御部が形成されたディスク制御ユニットと、

前記データを記憶するためのメモリが形成されたキャッシュメモリユニットと

前記ホストインタフェース制御部と前記ディスクインタフェース制御部と前記メモリとが形成されたストレージ制御ユニットと
を挿抜可能な装着部と、

前記チャネル制御ユニット、前記ディスク制御ユニット、前記キャッシュメモリユニット、及び前記ストレージ制御ユニットを通信可能に接続する内部接続部と
を備えることを特徴とする記憶デバイス制御装置。

【請求項 2】 前記ストレージ制御ユニットの前記メモリには、

前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリュームを特定するための情報が記憶され、

前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリュームを特定するための情報と他の前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリュームを特定するための情報との少なくともいずれかが記憶されること
を特徴とする請求項 1 に記載の記憶デバイス制御装置。

【請求項 3】 前記ストレージ制御ユニットの前記メモリは、

前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリュームに記憶される前記データを記憶するための第 1 の記憶領域と、

前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリュームに記憶される前記データを記憶するための第 2 の記憶領域とを有し、

前記第 1 及び前記第 2 の記憶領域を特定するための情報を記憶すること
を特徴とする請求項 1 に記載の記憶デバイス制御装置。

【請求項 4】 前記ストレージ制御ユニットは、前記他のストレージ制御ユニットとの間で授受されるデータを記憶するための通信バッファを備えることを特徴とする請求項 1 に記載の記憶デバイス制御装置。

【請求項 5】 データ入出力要求を受信するためのホストインタフェース制御部が形成されたチャンネル制御ユニットと、

前記データ入出力要求に応じて、データを記憶するための記憶ボリュームに対する前記データの入出力制御を行うためのディスクインタフェース制御部が形成されたディスク制御ユニットと、

前記データを記憶するためのメモリが形成されたキャッシュメモリユニットと

、

前記ホストインタフェース制御部と前記ディスクインタフェース制御部と前記メモリとが形成されたストレージ制御ユニットと

を挿抜可能な装着部と、

前記チャンネル制御ユニット、前記ディスク制御ユニット、前記キャッシュメモリユニット、及び前記ストレージ制御ユニットを通信可能に接続する内部接続部と

を備え、

前記装着部には少なくとも、

前記ストレージ制御ユニットの前記メモリに、

前記データ入出力要求の対象となっている前記記憶ボリュームに対する前記入出力制御を行う前記ユニットを特定するための情報が記憶された、複数の前記ストレージ制御ユニット

が装着される

記憶デバイス制御装置の制御方法であって、

ある前記ストレージ制御ユニットが、前記データ入出力要求を受信するステップと、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象とな

っている前記記憶ボリュームに対する前記入出力制御を行う前記ユニットを特定するステップと、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットである場合には、前記ストレージ制御ユニットが前記入出力制御を行い、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、他の前記ストレージ制御ユニットが前記入出力制御を行うステップと
を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 6】 前記ストレージ制御ユニットが前記受信する前記データ入出力要求が前記データの読み出し要求であって、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、

前記他のストレージ制御ユニットが前記入出力制御を行う前記ステップは、
前記ストレージ制御ユニットが、前記読み出し要求を前記他のストレージ制御ユニットに送信するステップと、

前記他のストレージ制御ユニットが、前記読み出し要求に応じて前記入出力制御を行うステップと、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットから前記データを受信するステップと、

前記ストレージ制御ユニットが、前記受信した前記データを情報処理装置に送信するステップと

であることを特徴とする請求項 5 に記載の記憶デバイス制御装置の制御方法。

【請求項 7】 前記ストレージ制御ユニットが前記受信する前記データ入出力要求が前記データの書き込み要求であって、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、

前記他のストレージ制御ユニットが前記入出力制御を行う前記ステップは、
前記ストレージ制御ユニットが、前記書き込み要求及び書き込みデータを前記他のストレージ制御ユニットに送信するステップと、

前記他のストレージ制御ユニットが、前記書き込み要求に応じて前記書き込みデータの前記入出力制御を行うステップと

であることを特徴とする請求項 5 に記載の記憶デバイス制御装置の制御方法。

【請求項 8】 請求項 7 に記載の記憶デバイス制御装置の制御方法において

、
前記ストレージ制御ユニットが、前記他のストレージ制御ユニットから前記書き込みデータの前記入出力制御を完了した旨のデータを受信するステップと、

前記ストレージ制御ユニットが、情報処理装置に前記入出力制御を完了した旨のデータを送信するステップと
を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 9】 前記ストレージ制御ユニットは、前記他のストレージ制御ユニットとの間で授受されるデータを記憶するための通信バッファを備えており、

前記ストレージ制御ユニットが、前記読み出し要求を前記他のストレージ制御ユニットに送信する前記ステップは、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットが備える前記通信バッファに前記読み出し要求を書き込むステップと、

前記他のストレージ制御ユニットが、前記他のストレージ制御ユニットの前記通信バッファから前記読み出し要求を読み出すステップとであり、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットから前記データを受信する前記ステップは、

前記他のストレージ制御ユニットにより前記ストレージ制御ユニットが備える前記通信バッファに書き込まれた前記データを、前記ストレージ制御ユニットが読み出すステップ

であることを特徴とする請求項 6 に記載の記憶デバイス制御装置の制御方法。

【請求項 1 0】 前記ストレージ制御ユニットは、前記他のストレージ制御ユニットとの間で授受されるデータを記憶するための通信バッファを備えており

、
前記ストレージ制御ユニットが、前記書き込み要求及び書き込みデータを前記他のストレージ制御ユニットに送信する前記ステップは、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットが備える前記通信バッファに前記書き込み要求及び前記書き込みデータを書き込むステップと、

前記他のストレージ制御ユニットが、前記他のストレージ制御ユニットの前記通信バッファから前記書き込み要求及び前記書き込みデータを読み出すステップと

であることを特徴とする請求項 7 に記載の記憶デバイス制御装置の制御方法。

【請求項 11】 請求項 8 に記載の記憶デバイス制御装置の制御方法において、

前記ストレージ制御ユニットは、前記他のストレージ制御ユニットとの間で授受されるデータを記憶するための通信バッファを備えており、

前記ストレージ制御ユニットが、前記他のストレージ制御ユニットから前記書き込みデータの前記入出力制御を完了した旨のデータを受信する前記ステップは、

前記他のストレージ制御ユニットにより前記ストレージ制御ユニットが備える前記通信バッファに書き込まれた前記データを、前記ストレージ制御ユニットが読み出すステップであること

を特徴とする記憶デバイス制御装置の制御方法。

【請求項 12】 前記装着部には前記ストレージ制御ユニットと前記ディスク制御ユニットとが少なくとも装着されている請求項 5 に記載の記憶デバイス制御装置の制御方法であって、

前記ストレージ制御ユニットが、前記データ入出力要求を受信するステップと、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象となっている前記記憶ボリュームに対する前記入出力制御を行う前記ユニットを特定するステップと、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットである場合には、前記ストレージ制御ユニットが前記入出力制御を行い、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、前記ディスク制御ユニットが前記入出力制御を行うステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 13】 前記ストレージ制御ユニットが前記受信する前記データ入

出力要求が前記データの読み出し要求であって、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、

前記ディスク制御ユニットが前記入出力制御を行う前記ステップは、
前記ストレージ制御ユニットが、前記読み出し要求を前記ディスク制御ユニットに送信するステップと、

前記ディスク制御ユニットが、前記読み出し要求に応じて前記入出力制御を行うステップと、

前記ストレージ制御ユニットが、前記ディスク制御ユニットから前記データを受信するステップと、

前記ストレージ制御ユニットが、前記受信した前記データを情報処理装置に送信するステップと

であることを特徴とする請求項 12 に記載の記憶デバイス制御装置の制御方法。

【請求項 14】 前記ストレージ制御ユニットが前記受信する前記データ入出力要求が前記データの書き込み要求であって、前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、

前記ディスク制御ユニットが前記入出力制御を行う前記ステップは、
前記ストレージ制御ユニットが、前記書き込み要求及び書き込みデータを前記ディスク制御ユニットに送信するステップと、

前記ディスク制御ユニットが、前記書き込み要求に応じて前記書き込みデータの前記入出力制御を行うステップと

であることを特徴とする請求項 12 に記載の記憶デバイス制御装置の制御方法。

【請求項 15】 請求項 12 に記載の記憶デバイス制御装置の制御方法であって、

前記ストレージ制御ユニットの前記メモリは、
前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリュームに記憶される前記データを記憶するための第 1 の記憶領域と、

前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリュームに記憶される前記データを記憶するための第 2 の記憶領域とを有し、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットである場合

には、前記ストレージ制御ユニットが、前記第1の記憶領域に対して前記データの入出力制御を行い、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットでない場合には、前記ストレージ制御ユニットが、前記第2の記憶領域に対して前記データの入出力制御を行うこと

を特徴とする記憶デバイス制御装置の制御方法。

【請求項16】 前記装着部には複数の前記ストレージ制御ユニットと前記ディスク制御ユニットとが少なくとも装着されている請求項5に記載の記憶デバイス制御装置の制御方法であって、

前記ストレージ制御ユニットが、前記データ入出力要求を受信するステップと

、
前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象となっている前記記憶ボリュームに対する前記入出力制御を行う前記ユニットを特定するステップと、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットである場合には、前記ストレージ制御ユニットが前記入出力制御を行い、前記入出力制御を行う前記ユニットが前記ディスク制御ユニットである場合には、前記ディスク制御ユニットが前記入出力制御を行い、前記入出力制御を行う前記ユニットが他の前記ストレージ制御ユニットである場合には、前記他のストレージ制御ユニットが前記入出力制御を行うステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項17】 前記装着部には前記ストレージ制御ユニットと前記ディスク制御ユニットと前記キャッシュメモリユニットとが少なくとも装着されている請求項5に記載の記憶デバイス制御装置の制御方法であって、

前記ストレージ制御ユニットの前記メモリは、

前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリュームに記憶される前記データを記憶するための第1の記憶領域と、

前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリュームに記憶される前記データを記憶するための第2の記憶領域とを有し、

前記ストレージ制御ユニットが、前記データ入出力要求を受信するステップと、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象となっている前記記憶ボリュームに対する前記入出力制御を行う前記ユニットを特定するステップと、

前記入出力制御を行う前記ユニットが前記ストレージ制御ユニットである場合には、前記ストレージ制御ユニットが、前記第 1 の記憶領域に対して前記データの入出力制御を行うステップと

前記データが前記第 1 の記憶領域に記憶されていない場合には、前記ストレージ制御ユニットは、前記キャッシュメモリユニットに対して前記データの入出力制御を行うステップと、

前記データが前記キャッシュメモリユニットに記憶されていない場合には、前記ストレージ制御ユニットが、前記記憶ボリュームに対して前記入出力制御を行うステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 1 8】 前記装着部には前記ストレージ制御ユニットと前記ディスク制御ユニットと前記キャッシュメモリユニットとが少なくとも装着されている請求項 5 に記載の記憶デバイス制御装置の制御方法であって、

前記ストレージ制御ユニットの前記メモリは、

前記ストレージ制御ユニットが前記入出力制御を行う前記記憶ボリュームに記憶される前記データを記憶するための第 1 の記憶領域と、

前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリュームに記憶される前記データを記憶するための第 2 の記憶領域とを有し、

前記ストレージ制御ユニットが、前記データ入出力要求を受信するステップと、

、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象となっている前記記憶ボリュームに対する前記入出力制御を行う前記ユニットを特定するステップと、

前記入出力制御を行う前記ユニットが前記ディスク制御ユニットである場合に

は、前記ストレージ制御ユニットが、前記第 2 の記憶領域に対して前記データの
入出力制御を行うステップと

前記データが前記第 2 の記憶領域に記憶されていない場合には、前記ストレ
ージ制御ユニットは、前記キャッシュメモリユニットに対して前記データの入出力
制御を行うステップと、

前記データが前記キャッシュメモリユニットに記憶されていない場合には、前
記ディスク制御ユニットが前記記憶ボリュームに対して前記入出力制御を行うステ
ップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【請求項 1 9】 請求項 5 に記載の記憶デバイス制御装置の制御方法におい
て、前記装着部に前記キャッシュメモリユニットが装着された場合には、

各前記ストレージ制御ユニットの前記メモリに記憶される、前記データ入出力
要求の対象となっている前記記憶ボリュームに対する前記入出力制御を行う前記ユ
ニットを特定するための情報の複製を、前記各ストレージ制御ユニットが、前記
キャッシュメモリユニットに書き込むステップ
を備え、

前記ストレージ制御ユニットが前記情報を参照し、前記入出力要求の対象とな
っている前記記憶ボリュームに対する前記入出力制御を行う前記ユニットを特定す
る前記ステップにおいて、

前記ストレージ制御ユニットが、前記ストレージ制御ユニットの前記メモリに
記憶されている前記情報を参照しても前記ユニットを特定できない場合には、前
記キャッシュメモリユニットの前記情報を参照して、前記ユニットを特定するこ
と

を特徴とする記憶デバイス制御装置の制御方法。

【請求項 2 0】 請求項 5 に記載の記憶デバイス制御装置の制御方法におい
て、前記装着部に前記ディスク制御ユニットが装着された場合には、

ある前記ストレージ制御ユニットが、前記あるストレージ制御ユニットにより
前記入出力制御が行われる前記記憶ボリュームに記憶される前記データの複製を、
前記ディスク制御ユニットが前記入出力制御を行う前記記憶ボリュームに、書き込

むステップと、

前記あるストレージ制御ユニット及び前記他のストレージ制御ユニットのそれぞれが各前記メモリに記憶している、前記入出力制御を行う前記ユニットを特定するための情報を、前記あるストレージ制御ユニットから前記ディスク制御ユニットに変更するステップと

を備えることを特徴とする記憶デバイス制御装置の制御方法。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、記憶デバイス制御装置、及び記憶デバイス制御装置の制御方法に関する。

【0 0 0 2】

【従来の技術】

コンピュータシステムにおいてデータの記憶装置として用いられるストレージシステムは、ユーザのニーズに応じて小規模コンピュータシステム向けのものから大規模コンピュータシステム向けのものまで、様々なものが提供されている。

【0 0 0 3】

小規模コンピュータシステム向けのストレージシステムは、ストレージシステムとしての一通りの機能を備えた装置として提供されることにより、導入の容易化、導入時のコスト抑制が図られている。

一方、大規模コンピュータシステム向けのストレージシステムは小規模向けストレージシステムとは異なるアーキテクチャが採用され、高い拡張性を備え、最大規模での運用を求めるユーザニーズにも応えることができる構成となっている。

【0 0 0 4】

【特許文献 1】

特開 2 0 0 2 - 1 2 3 4 7 9 号公報

【0 0 0 5】

【発明が解決しようとする課題】

しかしながら、記憶容量の増大や、ストレージシステムの統合等によるストレージシステムの大規模化が必要な場合、小規模コンピュータシステム向けストレージシステムを採用していた場合には、大規模向けストレージシステムに入れ替えるか、ストレージシステムを追加するなどの対応が必要となる。

一方、当初から高い拡張性を備えた大規模コンピュータシステム向けストレージシステムを採用していた場合には、コンピュータシステムの導入開始初期からストレージシステムに対して大きなコストの負担を強いられる。

本発明は上記課題を鑑みてなされたものであり、記憶デバイス制御装置、及び記憶デバイス制御装置の制御方法を提供することを主たる目的とする。

【0006】

【課題を解決するための手段】

上記課題を解決するために、本発明に係る記憶デバイス制御装置は、データ入出力要求を受信するためのホストインタフェース制御部が形成されたチャネル制御ユニットと、前記データ入出力要求に応じて、データを記憶するための記憶ボリュームに対する前記データの入出力制御を行うためのディスクインタフェース制御部が形成されたディスク制御ユニットと、前記データを記憶するためのメモリが形成されたキャッシュメモリユニットと、前記ホストインタフェース制御部と前記ディスクインタフェース制御部と前記メモリとが形成されたストレージ制御ユニットとを挿抜可能な装着部と、前記チャネル制御ユニット、前記ディスク制御ユニット、前記キャッシュメモリユニット、及び前記ストレージ制御ユニットを通信可能に接続する内部接続部とを備える。

【0007】

本発明に係る記憶デバイス制御装置においては、ストレージ制御ユニットや、チャネル制御ユニット、ディスク制御ユニット、グローバルキャッシュのいずれも装着することができるので、顧客ニーズに応じた柔軟性の高いストレージシステム構成することができる。

【0008】

また上記データ入出力要求とは例えばデータのリード要求やライト要求である。また入出力制御とは、データの読み出しや書き込みを行うための制御である。

記憶ボリュームとは、ハードディスク装置や半導体記憶装置等により構成される記憶装置により提供される物理的な記憶領域である物理ボリュームと、物理ボリューム上に論理的に設定される記憶領域である論理ボリュームとを含む記憶リソースである。

【0009】

その他、本願が開示する課題、及びその解決方法は、発明の実施の形態の欄、及び図面により明らかにされる。

【0010】

【発明の実施の形態】

以下、本発明の実施の形態について図面を用いて詳細に説明する。

===外観構成===

まず、本実施の形態に係るストレージシステム100の外観構成を示す図を図1に示す。

ストレージシステム100はディスク制御装置（記憶デバイス制御装置）110とディスク駆動装置120とを備えている。ディスク制御装置110はストレージシステム100全体の制御を司る。ディスク駆動装置120はデータを記憶するディスクドライブ121を多数収納する。図1に示すストレージシステム100ではディスク制御装置110が中央に配置され、その左右にディスク駆動装置120が配置されている。なお図1に示すように、ディスクドライブ121はディスク制御装置110にも収納されるようにすることができる。

【0011】

ディスク制御装置110は、コントローラ部111、ファン113、電源部112を備えている。コントローラ部111はストレージシステム100全体の制御を司る部分である。詳細は後述するが、コントローラ部111はチャネル制御ユニット300、ディスク制御ユニット400、ストレージ制御ユニット800、グローバルキャッシュ（キャッシュメモリユニット）600を含んで構成される。ディスク制御装置110にこれらのユニットが装着されることにより、ストレージシステム100の制御が行われる。これらのユニットは、後述するように、一体的にユニット化された回路基板上に形成されたハードウェアさらにこのハ

ードウェアにより実行されるソフトウェア又は両方により実現される。ファン 1 1 3 はディスク制御装置 1 1 0 を冷却するために用いられる。電源部 1 1 2 はディスク制御装置 1 1 0 への電力の供給を行うために用いられる。

【0 0 1 2】

ディスク駆動装置 1 2 0 には多数のディスクドライブ 1 2 1 が収納される。ディスクドライブ 1 2 1 は、ディスク駆動装置 1 2 0 を構成する筐体に着脱可能なように収納されている。

【0 0 1 3】

また図 1 には示されていないが、ディスク制御装置 1 1 0 には管理端末 1 6 0 が接続されている。管理端末 1 6 0 はストレージシステム 1 0 0 の保守管理を行うためのコンピュータである。管理端末 1 6 0 はストレージシステム 1 0 0 に組み込まれるように構成することもできるし、遠隔地に設置されストレージシステム 1 0 0 とネットワークで接続された形態とすることもできる。

【0 0 1 4】

===全体構成===

次に、本実施の形態に係るストレージシステムの全体構成を示すブロック図を図 2 に示す。

ディスク制御装置 1 1 0 はホスト計算機（情報処理装置） 2 0 0 と接続され、ホスト計算機 2 0 0 からのデータのリード／ライト要求（データ入出力要求）を受信する。また多数のディスクドライブ 1 2 1 と接続されており、ホスト計算機 2 0 0 からのデータの入出力要求に応じて、記憶ボリュームに対してデータの入出力制御を行う。記憶ボリュームとは、記憶装置により提供される物理的な記憶領域である物理ボリュームと、物理ボリューム上に論理的に設定される記憶領域である論理ボリューム 1 2 2 とを含む記憶リソースである。記憶装置としては、例えばハードディスク装置や半導体記憶装置等様々なものを採用することができる。

【0 0 1 5】

ディスク制御装置 1 1 0 とホスト計算機 2 0 0 との間の通信は、様々な通信プロトコルに従って行うようにすることができる。例えば、ファイバチャネルや S C S I （Small Computer System Interface）、F I C O N （Fibre Connection

) (登録商標)、E S C O N (Enterprise System Connection) (登録商標)、A C O N A R C (Advanced Connection Architecture) (登録商標)、F I B A R C (Fibre Connection Architecture) (登録商標)、T C P / I P (Transmission Control Protocol/Internet Protocol) 等である。これらの通信プロトコルを混在させるようにすることもできる。例えばホスト A 2 0 0 との間の通信はファイバチャネルで行い、ホスト B 2 0 0 との間の通信は T C P / I P で行うようにすることもできる。ホスト計算機 2 0 0 がメインフレーム計算機である場合には、例えば F I C O N や E S C O N、A C O N A R C、F I B A R C が用いられる。またホスト計算機 2 0 0 がオープン系の計算機である場合には、例えばファイバチャネルや S C S I、T C P / I P が用いられる。なお、ホスト計算機 2 0 0 からのデータのリード/ライト要求は、記憶ボリュームにおけるデータの管理単位であるブロックを単位として行うようにすることもできるし、ファイル名を指定することによりファイル単位に行うようにすることもできる。後者の場合にはディスク制御装置 1 1 0 は、ホスト計算機 2 0 0 からのファイルレベルでのアクセスを実現する N A S (Network Attached Storage) として機能する。

【 0 0 1 6 】

ホスト計算機 2 0 0 は C P U (Central Processing Unit) やメモリ、入出力装置等を備えたコンピュータである。ホスト計算機 2 0 0 には、図示されていないクライアント計算機が接続されている。ホスト計算機 2 0 0 はクライアント計算機に対して各種の情報処理サービスを提供する。ホスト計算機 2 0 0 により提供される情報処理サービスは、例えば銀行の自動預金預け払いサービスやインターネットのホームページ閲覧サービスのようなオンラインサービスを始め、科学技術分野における実験シミュレーションを行うバッチ処理サービス等である。またホスト計算機 2 0 0 とディスク制御装置 1 1 0 間のアクセスルートは 2 重化されており、片方のアクセスルートに障害が発生しても他方のアクセスルートにより入出力要求の受信を継続することができるようになっている。

【 0 0 1 7 】

図 2 に示すディスク制御装置 1 1 0 は、4 枚のストレージ制御ユニット 8 0 0、2 枚のチャネル制御ユニット 3 0 0、2 枚のディスク制御ユニット 4 0 0、2

枚のグローバルキャッシュ 600、及び内部接続部 500 を有している。またディスク制御装置 110 には管理端末 160 が接続されている。

【0018】

===ストレージ制御ユニット===

ストレージ制御ユニット 800 は、ホストインタフェース制御部（ホスト I/F 制御部）810、ディスクインタフェース制御部（ディスク I/F 制御部）860、キャッシュ制御部 820、ローカルキャッシュ（メモリ）830、内部インタフェース接続部（内部 I/F 接続部）840 を備える。ストレージ制御ユニット 800 は、これらが一体的にユニット化された回路基板上に形成されたハードウェアさらにこのハードウェアにより実行されるソフトウェア又は両方により実現される。

【0019】

ホスト I/F 制御部 810 は、ホスト計算機 200 とのインタフェース機能を有する。ディスク I/F 制御部 860 は、記憶ボリュームに対する入出力制御を行うためのインタフェース機能を備える。ローカルキャッシュ 830 は、ホスト計算機 200 と記憶ボリュームとの間で授受されるデータを記憶する。キャッシュ制御部 820 は、ローカルキャッシュ 830 の制御を司る。なお本実施の形態においては、ストレージ制御ユニット 800 は他のストレージ制御ユニット 800 との間でクラスタを構成している。クラスタを構成することにより、同一クラスタ内のあるストレージ制御ユニット 800 に障害が発生した場合でも、障害が発生したストレージ制御ユニット 800 がそれまで行っていた処理を同一クラスタ内の他のストレージ制御ユニット 800 に引き継ぐことにより、処理を継続することができる。キャッシュ制御部 820 は、クラスタを構成する他のストレージ制御ユニット 800 のキャッシュ制御部 820 とペア間接続部 850 を介して接続されている。クラスタを構成するストレージ制御ユニット 800 間でローカルキャッシュ 830 のデータを相互に記憶することによりデータの 2 重化を行っている。内部 I/F 制御部 840 は、内部接続部 500 を介してグローバルキャッシュ 600、ディスク制御ユニット 400、チャネル制御ユニット 300、他のストレージ制御ユニット 800 に接続されている。なおストレージ制御ユニット 800 は

、ホスト I F 制御部 810、ディスク I F 制御部 860、内部 I F 制御部 840 を備えて構成され、ローカルキャッシュ 830 及びキャッシュ制御部 820 を備えない構成とすることもできる。この場合、クラスタを構成する各ストレージ制御ユニット 800 は、例えば相互の内部 I F 制御部 840 をペア間接続部 850 により接続するようにすることができる。またホスト計算機 200 と記憶ボリューム間で授受されるデータは、ローカルキャッシュ 830 には記憶されずに、後述するグローバルキャッシュ 600 に記憶されるようにすることもできる。またローカルキャッシュ 830、グローバルキャッシュ 600 のいずれにも記憶されずに授受されるようにすることもできる。

【0020】

本実施の形態に係るストレージ制御ユニット 800 の外観構成を示す図を図 3 に示す。ストレージ制御ユニット 800 は、ディスク制御装置 110 が備える装着部 130 に挿入されることにより、ディスク制御装置 110 に装着される。図 7 にストレージ制御ユニット 800 がディスク制御装置 110 の装着部 130 に挿入される様子を示す。装着部 130 には複数のスロットが設けられており、各スロットにはストレージ制御ユニット 800 を装着するためのガイドレールが設けられている。ガイドレールに沿ってストレージ制御ユニット 800 をスロットに挿入することにより、ストレージ制御ユニット 800 をディスク制御装置 110 に装着することができる。各スロットに装着されたストレージ制御ユニット 800 は、ガイドレールに沿って引き抜くことにより取り外すことができる。またストレージ制御ユニット 800 には、ストレージ制御ユニット 800 とディスク制御装置 110 とを電氣的に接続するためのコネクタ 870 が設けられている。コネクタ 870 はディスク制御装置 110 の装着部 130 の奥手方向正面部に設けられた相手側コネクタと嵌合する。

【0021】

なおディスク制御装置 110 の各スロットには、ストレージ制御ユニット 800 のみならず、チャネル制御ユニット 300、ディスク制御ユニット 400、グローバルキャッシュ 600 も装着できる。いずれのユニットも、サイズやコネクタの位置、コネクタのピン配列等に互換性をもたせるようにしているからである

。従って、例えば全てのスロットにストレージ制御ユニット 8 0 0 を装着するようにすることもできるし、ディスク制御ユニット 4 0 0 やチャネル制御ユニット 3 0 0、グローバルキャッシュ 6 0 0 を混在させて装着するようにすることもできる。

【 0 0 2 2 】

上述のように、ストレージ制御ユニット 8 0 0 は、ホスト計算機 2 0 0 とのインタフェース機能を有するホスト I F 制御部 8 1 0 と、記憶ボリウムに対する入出力制御を行うためのインタフェース機能を備えるディスク I F 制御部 8 6 0 と、ホスト計算機 2 0 0 と記憶ボリウムとの間で授受されるデータを記憶するローカルキャッシュ 8 3 0 とが同一パッケージ上で構成されている。これにより、ストレージ制御ユニット 8 0 0 を追加していくことにより容易にシステムを拡張することが可能である。なおパッケージとは、複数の機能をモジュール化して 1 つの部品として構成したものである。部品の交換等の保守・管理はパッケージ単位に行われる。

【 0 0 2 3 】

またストレージ制御ユニット 8 0 0 において、ホスト I F 制御部 8 1 0 と、ディスク I F 制御部 8 6 0 と、ローカルキャッシュ 8 3 0 とが同一パッケージ上で構成されることにより、ホスト計算機 2 0 0 と記憶ボリウムとの間のデータ入出力性能を向上させることができる。

【 0 0 2 4 】

なぜならば、これらが同一パッケージ上で構成されることにより、ホスト計算機 2 0 0 と記憶ボリウムとの間のデータ転送路の電気的特性を向上させることができ、高速なデータ転送が可能となるからである。すなわち、ホスト I F 制御部 8 1 0、ディスク I F 制御部 8 6 0、ローカルキャッシュ 8 3 0 は、ホスト計算機 2 0 0 と記憶ボリウムとの間のデータ転送路の一部を構成するが、これらが同一パッケージ内に配置されることにより、データ転送路上に介在するコネクタやケーブルを削減できるため、例えばデータ転送路のインピーダンスを低下させ、耐ノイズ性を向上させることができるからである。さらに、ホスト I F 制御部 8 1 0、ディスク I F 制御部 8 6 0、ローカルキャッシュ 8 3 0 が同一パッケージ

内で近接して配置されることにより、これらを相互に接続する配線長を短くすることができる。これによって、ストレージ制御ユニット 8 0 0 内のデータ転送路のインピーダンスを低下させることができる。また耐ノイズ性も向上させることができる。これらにより、ストレージ制御ユニット 8 0 0 におけるデータ転送ピッチを上げることができるようになるため、ホスト計算機 2 0 0 と記憶ボリュームとの間のデータ入出力性能を向上させることができるようになる。

【 0 0 2 5 】

さらに、ストレージ制御ユニット 8 0 0 はホスト計算機 2 0 0 と記憶ボリュームとの両方に接続しているので、ホスト計算機 2 0 0 からのデータ入出力要求が自己に接続している記憶ボリュームに記憶されたデータに対するものである場合には、他のユニットを介さずに処理を行うことができる。このためホスト計算機 2 0 0 と記憶ボリュームとの間のデータ転送路がパッケージ間を跨るような処理が減少し、データ入出力性能を向上させることができるようになる。

【 0 0 2 6 】

またストレージ制御ユニット 8 0 0 は、ホスト計算機 2 0 0 と記憶ボリュームとの間のデータの授受をローカルキャッシュ 8 3 0 を使用せずに行うように制御することもできる。これにより、ローカルキャッシュ 8 3 0 を経由することによるデータ入出力処理の遅延を減少させることができる。これは、例えばホスト計算機 2 0 0 からのデータ入出力要求に応じて行われる記憶ボリュームへのデータアクセスに局所性がない場合などローカルキャッシュ 8 3 0 を用いても高いヒット率が期待できない場合に有効である。

【 0 0 2 7 】

また、ストレージ制御ユニット 8 0 0 を用いることにより、ホスト計算機 2 0 0 と記憶ボリュームとの間でパッケージ間を跨って行われるデータ転送が減少するので、万が一ストレージ制御ユニット 8 0 0 に障害が発生した場合でも、故障の影響を局所的に抑えることができる。つまりストレージ制御ユニット 8 0 0 に障害が発生した場合でも、他のストレージ制御ユニット 8 0 0 を用いて行われるデータ転送へ与える影響を少なくすることができる。同様に、例えば保守作業のためにストレージ制御ユニット 8 0 0 の交換を行う場合にも、その影響を局所的に

抑えることができ、他のストレージ制御ユニット 8 0 0 を用いて行われるデータ転送へ与える影響を少なくすることができる。

【0 0 2 8】

ストレージ制御ユニット 8 0 0 を用いたディスク制御装置 1 1 0 は、初期導入時点でのコスト効果が最大になるようにしたものであり、なおかつ拡張性を維持しているため小中規模から大規模構成向きである。例えば後述するように、ストレージ制御ユニット 8 0 0 を、電源 1 1 2 やファン 1 1 3 等と共に一つの筐体内に格納し、モジュラー型コントローラ 1 1 1 として構成するようにすることもできる。この場合にはストレージシステム 1 0 0 の初期導入を容易に行うことができる。またシステムの拡張時もモジュラー型コントローラ 1 1 1 を順次増設することにより、容易に行うことができる。従って、これからビジネスを開始する顧客やビジネス環境の変化が激しい顧客が状況に応じてシステム規模を変更できる柔軟性のあるシステムを実現するときに有効である。またディスク I F 制御部 8 6 0 が記憶ボリュームから読み出したデータはキャッシュ制御部 8 2 0 を経由してローカルキャッシュ 8 3 0 のデータ領域 8 3 1 に格納される。内部接続部 5 0 0 やグローバルキャッシュ 6 0 0 を経由しないので、高速なデータの読み出しが可能である。

【0 0 2 9】

一方、チャネル制御ユニット 3 0 0 とディスク制御ユニット 4 0 0 を用いたディスク制御装置 1 1 0 は、システムの最大規模を想定して最大規模でのコスト低減効果が最大になるようにしたものであり、大規模構成向きである。従って既にビジネスが安定している顧客が比較的大規模な構成を実現するときに有効である。

【0 0 3 0】

上述のように本実施の形態に係るディスク制御装置 1 1 0 においては、ストレージ制御ユニット 8 0 0 や、チャネル制御ユニット 3 0 0、ディスク制御ユニット 4 0 0、グローバルキャッシュ 6 0 0 のいずれも装着することができるので、顧客ニーズに応じた柔軟性の高いストレージシステム 1 0 0 を構成することができる。

【0031】

次にストレージ制御ユニット 800 の構成を示すブロック図を図 8 に示す。

ホスト I F 制御部 810 は、プロセッサ 811、メモリ 812、ホスト I F 回路 814、及び内部接続 I F 回路 815 を備えている。プロセッサ 811 はメモリ 812 に記憶されている制御プログラム 813 を実行することにより、ホスト 計算機 200 とのインタフェース機能を実現する。ホスト I F 回路 814 はホスト 計算機 200 と接続され、データの授受を行うための回路を構成する。内部接続 I F 回路 815 はキャッシュ制御部 820 との接続のための回路を構成する。

【0032】

キャッシュ制御部 820 は、キャッシュ制御部 I F 回路 821、バッファメモリ 822、内部接続 I F 回路 823、824、ペア間接続 I F 回路 825 を備える。キャッシュ制御部 I F 回路 821 は、ローカルキャッシュ 830 と接続するための回路を構成し、ローカルキャッシュ 830 との間のデータの授受を制御する。バッファメモリ 822 は、ローカルキャッシュ 830 との間のデータ授受の際に一時的にデータを格納するために用いられる。内部接続 I F 回路 823 は、ホスト I F 制御部 810、ディスク I F 制御部 860 と接続するための回路を構成する。内部接続 I F 回路 824 は、内部 I F 制御部 840 と接続するための回路を構成する。

【0033】

ペア間接続 I F 回路 825 は、クラスタを構成する相手のストレージ制御ユニット 800 のキャッシュ制御部 820 と接続するための回路を構成する。接続の様子を図 9 に示す。

クラスタを構成するストレージ制御ユニット 800 は相互にローカルキャッシュ 330 のデータを共有することによりデータの 2 重化を行っている。データの 2 重化を行うためのコマンドやデータはペア間接続 I F 回路 825 を介して相手側のストレージ制御ユニット 800 に送られる。相互のペア間接続 I F 回路 825 の間はペア間接続部 850 により直結されている。ペア間接続部 850 は相互のローカルキャッシュ 830 のデータを 2 重化するために設けられた通信路である。なおペア間接続部 850 はデータの 2 重化の他、例えばクラスタを構成する

ストレージ制御ユニット 800 間でのメッセージ通信を行うために使用することもできる。またハートビート信号を授受するために使用するようにすることもできる。ここでハートビート信号とは、クラスタを構成する各ストレージ制御ユニット 800 が相互に相手の動作状態を確認し合うための信号である。

【0034】

ディスク I/F 制御部 860 は、プロセッサ 861、メモリ 862、ディスク I/F 回路 864、及び内部接続 I/F 回路 865 を備えている。プロセッサ 861 はメモリ 862 に記憶されている制御プログラム 863 を実行することにより、ディスクドライブ 121 とのインタフェース機能を実現する。ディスク I/F 回路 864 はディスクドライブ 121 と接続され、データの授受を行うための回路を構成する。内部接続 I/F 回路 865 はキャッシュ制御部 820 との接続のための回路を構成する。

【0035】

なお、内部接続 I/F 回路 815、865、内部接続 I/F 回路 823、824、及びペア間接続 I/F 回路 825 の回路構成は、同種、異種、もしくは同種・異種の混在のいずれの態様とすることもできる。

【0036】

ローカルキャッシュ 830 は、データ領域 831 と制御領域 832 を有している。データ領域 831 は、ホスト計算機 200 と記憶ボリュームとの間で授受されるデータを記憶するための記憶領域である。制御領域 832 は、データ領域 831 に記憶されているデータを管理するための記憶領域である。ローカルキャッシュ 830 の詳細については後述する。

【0037】

===チャネル制御ユニット===

次にチャネル制御ユニット 300 の構成を示すブロック図を図 10 及び図 11 に示す。またチャネル制御ユニット 300 の外観構成を示す図を図 4 に示す。

チャネル制御ユニット 300 は、ホストインタフェース制御部（ホスト I/F 制御部）310、キャッシュ制御部 320、ローカルキャッシュ（メモリ）330、内部インタフェース接続部（内部 I/F 接続部）340 を備える。チャネル制御

ユニット 300 は、これらが一体的にユニット化された回路基板上に形成されたハードウェアさらにこのハードウェアにより実行されるソフトウェア又は両方により実現される。

【0038】

ホスト I/F 制御部 310 は、ホスト計算機 200 とのインタフェース機能を有する。ホスト I/F 制御部 310 は、プロセッサ 311、メモリ 312、ホスト I/F 回路 314、及び内部接続 I/F 回路 315 を備えている。ホスト I/F 制御部 310 により実現される機能、及び構成等はホスト I/F 制御部 810 と同様である。

【0039】

キャッシュ制御部 320、ローカルキャッシュ 330 は、ホスト計算機 200 と記憶ボリュームとの間で授受されるデータを記憶する。キャッシュ制御部 320、ローカルキャッシュ 330 により実現される機能、及び構成等もストレージ制御ユニット 800 におけるキャッシュ制御部 820、ローカルキャッシュ 830 と同様である。

内部 I/F 制御部 340 により実現される機能、構成等についても、ストレージ制御ユニット 800 における内部 I/F 制御部 840 と同様である。

【0040】

またチャネル制御ユニット 300 は、ストレージ制御ユニット 800 と同様ディスク制御装置 110 が備える装着部 130 に設けられたスロットに挿入することにより、ディスク制御装置 110 に装着される。チャネル制御ユニット 300 がディスク制御装置 110 の装着部 130 に挿入される様子を図 7 に示す。チャネル制御ユニット 300 には、チャネル制御ユニット 300 とディスク制御装置 110 とを電氣的に接続するためのコネクタ 370 が設けられている。コネクタ 370 はディスク制御装置 110 の装着部 130 の奥手方向正面部に設けられた相手側コネクタと嵌合する。前述したように、チャネル制御ユニット 300 は他のユニットとサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせるようにしている。そのため、ディスク制御装置 110 の各スロットには、ストレージ制御ユニット 800、チャネル制御ユニット 300、ディスク制御ユニッ

ト 4 0 0、グローバルキャッシュ 6 0 0 を混在させて装着するようにできる。

【 0 0 4 1 】

=== ディスク制御ユニット ===

次に、ディスク制御ユニット 4 0 0 の構成を示すブロック図を図 1 2 に示す。
またディスク制御ユニット 4 0 0 の外観構成を示す図を図 5 に示す。

ディスク制御ユニット 4 0 0 は、ディスクインタフェース制御部（ディスク I F 制御部） 4 6 0、内部インタフェース接続部（内部 I F 接続部） 4 4 0 を備える。ディスク制御ユニット 4 0 0 は、これらが一体的にユニット化された回路基板上に形成されたハードウェアさらにこのハードウェアにより実行されるソフトウェアまたは両方により実現される。

【 0 0 4 2 】

ディスク I F 制御部 4 6 0 は、ディスクドライブ 1 2 1 に対する入出力制御を行うためのインタフェース機能を有する。ディスク I F 制御部 4 6 0 は、プロセッサ 4 6 1、メモリ 4 6 2、ディスク I F 回路 4 6 4、及び内部接続 I F 回路 4 6 5 を備えている。ディスク I F 制御部 4 6 0 により実現される機能、及び構成等はストレージ制御ユニット 8 0 0 におけるディスク I F 制御部 8 6 0 と同様である。

内部 I F 制御部 4 4 0 により実現される機能、構成等についても、ストレージ制御ユニット 8 0 0 における内部 I F 制御部 8 4 0 と同様である。

【 0 0 4 3 】

また、ディスク制御ユニット 4 0 0 は、ストレージ制御ユニット 8 0 0 と同様ディスク制御装置 1 1 0 が備える装着部 1 3 0 に設けられたスロットに挿入することにより、ディスク制御装置 1 1 0 に装着される。ディスク制御ユニット 4 0 0 がディスク制御装置 1 1 0 の装着部 1 3 0 に挿入される様子を図 7 に示す。ディスク制御ユニット 4 0 0 には、ディスク制御ユニット 4 0 0 とディスク制御装置 1 1 0 とを電氣的に接続するためのコネクタ 4 7 0 が設けられている。コネクタ 4 7 0 はディスク制御装置 1 1 0 の装着部 1 3 0 の奥手方向正面部に設けられた相手側コネクタと嵌合する。前述したように、ディスク制御ユニット 4 0 0 は他のユニットとサイズやコネクタの位置、コネクタのピン配列等に互換性をもた

せるようにしている。そのため、ディスク制御装置 1 1 0 の各スロットには、ストレージ制御ユニット 8 0 0、チャネル制御ユニット 3 0 0、ディスク制御ユニット 4 0 0、グローバルキャッシュ 6 0 0 を混在させて装着するようにできる。

【 0 0 4 4 】

=== ローカルキャッシュ ===

次に、ストレージ制御ユニット 8 0 0 が備えるローカルキャッシュ 8 3 0 について図 1 3 を用いて説明する。なおチャネル制御ユニット 3 0 0 が備えるローカルキャッシュ 3 3 0 についても、その機能、構成等はストレージ制御ユニット 8 0 0 が備えるローカルキャッシュ 8 3 0 と同様である。

【 0 0 4 5 】

ローカルキャッシュ 8 3 0 は、データ領域 8 3 1 と制御領域 8 3 2 とを有している。データ領域 8 3 1 は、ホスト計算機 2 0 0 と記憶ボリュームとの間で授受されるデータを記憶するための記憶領域である。制御領域 8 3 2 は、データ領域 8 3 1 に記憶されているデータを管理するための記憶領域である。

【 0 0 4 6 】

データ領域 8 3 1 は、ダイレクトアクセス用データ領域 8 3 6 と、通信バッファ 8 3 7 を有している。ダイレクトアクセス用データ領域 8 3 6 はさらに、自 S A V O L (Storage Adapter V O L u m e) 用領域 (第 1 の記憶領域) 8 3 6 A と、他 D A V O L (Disk Adapter V O L u m e) 用領域 (第 2 の記憶領域) 8 3 6 B とを有している。

【 0 0 4 7 】

自 S A V O L 用領域 8 3 6 A は、ホスト計算機 2 0 0 から受信したデータ入出力要求の対象となっている記憶ボリュームが、当該データ入出力要求を受信したストレージ制御ユニット 8 0 0 に接続された記憶ボリュームに対するものである場合に、ホスト計算機 2 0 0 と当該記憶ボリュームとの間で授受されるデータを記憶するための領域である。

【 0 0 4 8 】

他 D A V O L 用領域 8 3 6 B は、ホスト計算機 2 0 0 から受信したデータ入出力要求の対象となっている記憶ボリュームが、ディスク制御ユニット 4 0 0 に接続

された記憶ボリュームに対するものである場合に、ホスト計算機 2 0 0 と当該記憶ボリュームとの間で授受されるデータを記憶するための領域である。他 D A V O L 用領域 8 3 6 B は、ディスク制御装置 1 1 0 にディスク制御ユニット 4 0 0 が装着された場合に設けられる記憶領域である。

【 0 0 4 9 】

通信バッファ 8 3 7 は、ホスト計算機 2 0 0 から受信したデータ入出力要求の対象となっている記憶ボリュームが、当該データ入出力要求を受信したストレージ制御ユニット 8 0 0 とは別のストレージ制御ユニット 8 0 0 に接続された記憶ボリュームに対するものである場合に、当該別のストレージ制御ユニット 8 0 0 との間でデータ入出力要求及びデータを授受するための記憶領域である。通信バッファ 8 3 7 は、ディスク制御装置 1 1 0 に異なるクラスタに属する複数のストレージ制御ユニット 8 0 0 が装着された場合に設けられる記憶領域である。

【 0 0 5 0 】

制御領域 8 3 2 には、キャッシュ領域管理テーブル 8 3 3、キャッシュデータ管理テーブル 8 3 4、ボリューム管理テーブル 8 3 5 が記憶されている。図 1 3 に示した例では、一つのキャッシュ領域管理テーブル 8 3 3 と、2つのキャッシュデータ管理テーブル 8 3 4 A、8 3 4 B と、一つのボリューム管理テーブル 8 3 5 とが記載されているが、これらのテーブルは適宜複数に分割されているようにすることもできる。

【 0 0 5 1 】

キャッシュ領域管理テーブル 8 3 3 は、データ領域 8 3 1 に設けられる自 S A V O L 用領域 8 3 6 A、他 D A V O L 用領域 8 3 6 B、及び通信バッファ 8 3 7 のそれぞれの記憶領域を特定するための情報を記憶したテーブルである。記憶領域を特定するための情報は例えばローカルキャッシュのアドレス情報である。図 1 3 の例では、データ領域 8 3 1 に設定されたアドレスのうち、“ 0 0 0 0 0 0 0 0 番地” から “ A F F F F F F F 番地” まだが自 S A V O L 用領域 8 3 6 A であり、“ B 0 0 0 0 0 0 0 0 番地” から “ E F F F F F F F 番地” まだが他 D A V O L 用領域 8 3 6 B であり、“ F 0 0 0 0 0 0 0 0 番地” から “ F F F F F F F F 番地” まだが通信バッファ 8 3 7 であることが示されている。キャッシュ領域管

理テーブル 833 の内容を変更することにより、上記各領域の割り当てを変更することができる。例えば、ホスト計算機 200 から受信するデータ入出力要求の多くが、当該データ入出力要求を受信したストレージ制御ユニット 800 に接続された記憶ボリュームに対するものである場合には、自 SAVOL 用領域 836A の割り当てを増やすようにすることができる。これによりホスト計算機 200 からのデータ入出力要求に対するローカルキャッシュ 830 のキャッシュヒット率が上がることが期待できるので、ストレージシステム 100 の性能向上を図ることができる。キャッシュ領域管理テーブル 833 の内容の変更は、例えばストレージシステム 100 の保守管理を行うオペレータにより、管理端末 160 から行うようにすることができる。

【0052】

キャッシュデータ管理テーブル 834 は、データ領域 831 に記憶されているデータを管理するためのテーブルである。キャッシュデータ管理テーブル 834 は、データのデータブロック毎に、“Valid”、“Dirty”、“Address”、“Lock”、“Owner”、“Pointer” の欄を有する。

【0053】

なおデータ領域 831 に記憶されるデータブロックは、どのような単位で記憶されるようにすることも可能である。ディスクドライブ 121 のブロック単位やシリンダの単位、あるいはトラックの単位等に限定されるものではない。またデータブロックのサイズは可変長とすることもできるし、固定長とすることもできる。

【0054】

“Valid” 欄は、当該データブロックのデータが有効か否かを示す。ホスト計算機 200 からデータの読み出し要求があった場合に、当該データをデータ領域 831 に見つけることができても、当該データが有効でなければキャッシュアクセスはミスヒットとなる。

“Dirty” 欄は、記憶ボリュームからローカルキャッシュ 830 に読み出されたデータがホスト計算機 200 により書き換えられているか否かを示す。書き

換えられている場合は、当該データをディスクドライブ 1 2 1 へ書き戻しておく必要がある。書き換えられていなければ、当該データをディスクドライブ 1 2 1 へ書き戻す必要はない。

” A d d r e s s ” 欄は、ローカルキャッシュ 8 3 0 に記憶されるデータの記憶位置を示す。

【 0 0 5 5 】

” L o c k ” 欄は、クラスタを構成するストレージ制御ユニット 8 0 0 のローカルキャッシュ 8 3 0 間で相互に記憶されている当該データに対する処理を禁止するか否かを示すための欄である。ローカルキャッシュ 8 3 0 は、クラスタを構成する相手のローカルキャッシュ 8 3 0 とデータ 2 重化用の通信路であるペア間接続部 8 5 0 で接続されており、一方のローカルキャッシュ 8 3 0 に記憶されているデータが更新された場合は、ペアを組んでいるもう一方のローカルキャッシュ 8 3 0 にも 2 重化して記憶される。しかし 2 重化を完全同時に行うことはできないため、短時間ではあるが相互のローカルキャッシュ 8 3 0 に記憶されるデータに不一致が生じる。データが不一致の間に、例えば片方のローカルキャッシュ 8 3 0 から当該データがリプレイス（グローバルキャッシュ 6 0 0 やディスクドライブ 1 2 1 への書き戻し）されてしまうと、誤ったデータがグローバルキャッシュ 6 0 0 やディスクドライブ 1 2 1 に記憶されることも起こりうる。このような問題を発生させないために” L o c k ” 欄が設けられ、L o c k が有効な間は当該データに対する更新やリプレイス等の制御は禁止される。

【 0 0 5 6 】

” O w n e r ” 欄は、ペアを組むローカルキャッシュ 8 3 0 のどちらが当該データを所有しているのかを示す。ペア間では相互にデータを 2 重化して記憶し合っているため、どちらのデータであるのかを管理するために” O w n e r ” 欄が設けられている。

” P o i n t e r ” 欄は、データ領域 8 3 1 に記憶されるデータと制御領域 8 3 2 に記憶されるキャッシュデータ管理テーブル 8 3 4 の対応付けを管理するための欄である。

【 0 0 5 7 】

ボリウム管理テーブル 835 は、ホスト計算機 200 からのデータ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行うユニットを特定するための情報を記憶するためのテーブルである。ボリウム管理テーブル 835 は、“CA No” 欄、“path No” 欄、“DA No” 欄、“Volume No” 欄、“drive No” 欄、“config” 欄、“Access Method” 欄を有する。

【0058】

“CA No” 欄は、ディスク制御装置 110 に装着されているストレージ制御ユニット 800 のホスト I/F 制御部 810、またはチャネル制御ユニット 300 のホスト I/F 制御部 310 に付与された識別番号を記憶するための欄である。図 13 の例では、CA00 及び CA01 が “CA No” 欄に記載されている。CA00 及び CA01 は、図 2 において示されるように、相互にクラスタを組むストレージ制御ユニット 800 のホスト I/F 制御部 810 である。

【0059】

“path No” 欄は、ホスト計算機 200 からアクセス可能な論理ボリウム 122 を特定したパスに対して付与される識別番号を記憶するための欄である。本実施の形態においては、パスはストレージ制御ユニット 800 やチャネル制御ユニット 300 毎に付与される。従って同じパス番号でもユニットが異なれば別のパスである。パス番号をストレージシステム 100 全体でユニークな番号として付与するようにすることもできる。

【0060】

“DA No” 欄は、ディスク制御装置 110 に装着されているストレージ制御ユニット 800 のディスク I/F 制御部 860、またはディスク制御ユニット 400 のディスク I/F 制御部 460 に付与された識別番号を記憶するための欄である。図 13 の例では、DA00 及び DA01、DA02 及び DA03、DA04 及び DA05 が “DA No” 欄に記載されている。DA00 及び DA01 は、CA00 及び CA01 で識別されるホスト I/F 制御部 810 を含むストレージ制御ユニット 800 と同一のストレージ制御ユニット 800 のディスク I/F 制御部 860 である。DA02 及び DA03 は、CA00 及び CA01 で識別されるホ

スト I F 制御部 810 を含むストレージ制御ユニット 800 とは異なるストレージ制御ユニット 800 のディスク I F 制御部 860 である。DA04 及び DA05 は、CA00 及び CA01 で識別されるホスト I F 制御部 810 を含むストレージ制御ユニット 800 とは異なるディスク制御ユニット 400 のディスク I F 制御部 460 である。このように、CA00、CA01 で受信したホスト計算機 200 からのデータ入出力要求は、自己のストレージ制御ユニット 800 のディスク I F 制御部 860 に接続される記憶ボリュームに対してのみならず、他のストレージ制御ユニット 800 やディスク制御ユニット 400 のディスク I F 制御部 860、460 に接続される記憶ボリュームに対しても行われる。

【0061】

” Volume No” 欄は、” DA No” 欄で指定されるディスク I F 制御部 860、460 に接続されている論理ボリューム 122 を特定するための欄である。

” drive No” 欄は、” DA No” 欄で指定されるディスク I F 制御部 860、460 に接続されているディスクドライブ 121 を特定するための欄である。

” config” 欄は、” drive No” 欄で特定されるディスクドライブ 121 上に設定されている RAID (Redundant Arrays of Inexpensive Disks) の構成を記憶するための欄である。

【0062】

” Access Method” 欄は、ホスト計算機 200 から受信したデータ入出力要求の対象となっている記憶ボリュームに対する入出力制御を行う方法を特定するための欄である。「direct」と記載されている場合は、ホスト計算機 200 から受信したデータ入出力要求に指定されている当該データの記憶アドレスに基づき、当該データの入出力制御を行う。「message」と記載されている場合は、” DA No” 欄で特定されるディスク I F 制御部 860、460 を含むストレージ制御ユニット 800 またはディスク制御ユニット 400 に対して、ホスト計算機 200 から受信したデータ入出力要求を送信する。そして当該ストレージ制御ユニット 800 またはディスク制御ユニット 400 により入出

力制御が行われる。

【0 0 6 3】

このようにボリウム管理テーブル 8 3 5 を用いることにより、異種ユニットが混在して装着されているディスク制御装置 1 1 0 においても、ホスト計算機 2 0 0 からのデータ入出力要求に対するデータ入出力制御を行うことができる。

なお、ボリウム管理テーブル 8 3 5 にはディスクドライブ 1 2 1 の領域を指定するためのアドレス情報を記憶するための欄を設けるようにすることもできる。

【0 0 6 4】

=== グローバルキャッシュ ===

次に、グローバルキャッシュ 6 0 0 の構成を示すブロック図を図 1 4 に示す。またグローバルキャッシュ 6 0 0 の外観構成を示す図を図 6 に示す。

グローバルキャッシュ 6 0 0 は、ストレージ制御ユニット 8 0 0 やチャンネル制御ユニット 3 0 0、ディスク制御ユニット 4 0 0 と同様ディスク制御装置 1 1 0 が備える装着部 1 3 0 に設けられたスロットに挿入することにより、ディスク制御装置 1 1 0 に装着される。グローバルキャッシュ 6 0 0 がディスク制御装置 1 1 0 の装着部 1 3 0 に挿入される様子を図 7 に示す。グローバルキャッシュ 6 0 0 には、グローバルキャッシュ 6 0 0 とディスク制御装置 1 1 0 とを電氣的に接続するためのコネクタ 6 7 0 が設けられている。コネクタ 6 7 0 はディスク制御装置 1 1 0 の装着部 1 3 0 の奥手方向正面部に設けられた相手側コネクタと嵌合する。前述したように、グローバルキャッシュ 6 0 0 は他のユニットとサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせるようにしている。そのため、ディスク制御装置 1 1 0 の各スロットには、ストレージ制御ユニット 8 0 0、チャンネル制御ユニット 3 0 0、ディスク制御ユニット 4 0 0、グローバルキャッシュ 6 0 0 を混在させて装着するようにできる。

【0 0 6 5】

グローバルキャッシュ 6 0 0 は、データ領域 6 0 1 と制御領域 6 0 2 とを有している。データ領域 6 0 1 は、ホスト計算機 2 0 0 と記憶ボリウムとの間で授受されるデータを記憶するための記憶領域である。制御領域 6 0 2 は、データ領域 6 0 1 に記憶されているデータを管理するための記憶領域である。

データ領域 601 は、ダイレクトアクセス用データ領域 606 と、通信バッファ 607 を有している。

ダイレクトアクセス用データ領域 606 は、ホスト計算機 200 と当該記憶ボリュームとの間で授受されるデータを記憶するための領域である。

【0066】

通信バッファ 607 は、ストレージ制御ユニット 800 間でデータ入出力要求及びデータを授受する際に使用される記憶領域である。なおストレージ制御ユニット 800 とディスク制御ユニット 400 との間でデータ入出力要求及びデータを授受する際に使用するようにすることもできる。またローカルキャッシュ 830 に通信バッファ 837 が設けられる場合には、グローバルキャッシュ 600 に通信バッファ 607 を設けないようにすることもできる。逆にグローバルキャッシュ 600 に通信バッファ 607 が設けられる場合には、ローカルキャッシュ 830 に通信バッファ 837 を設けないようにすることもできる。

【0067】

制御領域 602 には、キャッシュ領域管理テーブル 603、キャッシュデータ管理テーブル 604、ボリューム管理テーブル 605 が記憶されている。図 14 に示す例では、一つのキャッシュ領域管理テーブル 603 と、一つのキャッシュデータ管理テーブル 604 と、一つのボリューム管理テーブル 605 とが記載されているが、これらのテーブルは適宜複数に分割されているようにすることもできる。

【0068】

キャッシュ領域管理テーブル 603 は、データ領域 601 に設けられるダイレクトアクセス用データ領域 606、及び通信バッファ 607 のそれぞれの記憶領域を特定するための情報を記憶したテーブルである。記憶領域を特定するための情報は例えばグローバルキャッシュ 600 のアドレス情報である。図 14 の例では、データ領域 601 に設定されたアドレスのうち、“00000000番地”から“FFFFFFF番地”までがダイレクトアクセス用データ領域 606 であり、“F0000000番地”から“FFFFFFFF番地”までが通信バッファ 607 であることが示されている。キャッシュ領域管理テーブル 603 の内

容の変更は、例えばストレージシステム 1 0 0 の保守管理を行うオペレータにより、管理端末 1 6 0 から行うようにすることができる。これによりホスト計算機 2 0 0 からのデータ入出力要求の特性に応じたグローバルキャッシュ 6 0 0 設定が行えるので、ストレージシステム 1 0 0 の性能向上を図ることができる。

【 0 0 6 9 】

キャッシュデータ管理テーブル 6 0 4 は、データ領域 6 0 1 に記憶されているデータを管理するためのテーブルである。キャッシュデータ管理テーブル 6 0 4 の基本的な構成はローカルキャッシュ 8 3 0 のキャッシュデータ管理テーブル 8 3 4 の場合と同様であるが、“ L o c k ” 欄と“ O w n e r ” 欄の表す意味が異なる。

【 0 0 7 0 】

“ L o c k ” 欄は、グローバルキャッシュ 6 0 0 上の当該データがローカルキャッシュ 8 3 0 に読み出されており、ホスト計算機 2 0 0 により更新される可能性があるため、他のローカルキャッシュ 8 3 0 への読み出しは禁止されている状態であることを表す。複数のローカルキャッシュ 8 3 0 にデータの読み出しを許してしまうと、それぞれ独立にホスト計算機 2 0 0 によって更新される可能性があり、データの一致性が保証できなくなるためである。

“ O w n e r ” 欄は、当該データを読み出し中のローカルキャッシュ 8 3 0 を表す。

【 0 0 7 1 】

グローバルキャッシュ 6 0 0 は内部接続部 5 0 0 に接続されており、2 つのグローバルキャッシュ 6 0 0 がペアとなってデータの 2 重化を行っている。グローバルキャッシュ 6 0 0 間のデータの 2 重化は、内部接続部 5 0 0 を介して相互にデータを転送することにより実現される。

グローバルキャッシュ 6 0 0 におけるボリウム管理テーブル 6 0 5 は、ローカルキャッシュ 8 3 0 のボリウム管理テーブル 8 3 5 が複製されたものである。ローカルキャッシュ 8 3 0 が複数ある場合には、各ローカルキャッシュ 8 3 0 のボリウム管理テーブルの複製の結合である。

【 0 0 7 2 】

これにより、例えばあるストレージ制御ユニット 8 0 0 がホスト計算機 2 0 0 からデータ入出力要求を受信した場合に、当該ストレージ制御ユニット 8 0 0 のローカルキャッシュ 8 3 0 を参照しても、データ入出力要求の対象となっている記憶ボリュームに対するユニットが特定できなかった場合に、グローバルキャッシュ 6 0 0 のボリューム管理テーブル 6 0 5 を参照することにより記憶ボリュームに対する入出力制御を行うユニットを特定することができる。

【 0 0 7 3 】

===内部接続部===

次に、本実施の形態に係る内部接続部 5 0 0 の構成を示すブロック図を図 1 5 に示す。

内部接続部 5 0 0 は、ストレージ制御ユニット 8 0 0、チャネル制御ユニット 3 0 0、ディスク制御ユニット 4 0 0、グローバルキャッシュ 6 0 0 を相互に結合するためのスイッチである。

図 1 5 には 4 入力 4 出力の場合を示したが、実際の入出力数はディスク制御装置 1 1 0 に装着可能なユニットの数に応じた数となる。

【 0 0 7 4 】

内部接続部 5 0 0 は、受信部 5 1 0、送信部 5 2 0、制御部 5 3 0 を備えている。受信部 5 1 0 は、内部接続部 5 0 0 に入力されてきたデータを適宜バッファ 5 1 1 に蓄えつつ、制御部 5 3 0 からの指令に従って、指定された送信部 5 2 0 のバッファ 5 2 1 へデータを転送する。送信部 5 2 0 はバッファ 5 2 1 に格納されたデータを順次出力する。図 1 5 では内部接続部 5 0 0 としてクロスバスイッチの構成をとった場合を示したが、クロスバスイッチの構成に限られることはなく、様々な構成をとることが可能である。例えば受信部 5 1 0 と送信部 5 2 0 の間を多段のスイッチ回路で接続するようにすることもできる。

【 0 0 7 5 】

===管理端末===

次に、本実施の形態に係る管理端末 1 6 0 の構成を示すブロック図を図 1 6 に示す。

管理端末 1 6 0 は、CPU 1 6 1、メモリ 1 6 2、ポート 1 6 3、記録媒体読

取装置 164、入力装置 165、出力装置 166、記憶装置 168 を備える。

【0076】

CPU 161 は管理端末 160 の全体の制御を司るもので、記憶装置 168 に格納された管理プログラム 169 を適宜メモリ 162 に読み出して実行することによりストレージシステム 100 の保守管理のための各種機能を実現する。例えばディスクドライブ 121 上への論理ボリューム 122 の設定や、ストレージ制御ユニット 800 のホスト IF 制御部 810 において実行されるプログラム 813 のインストール等を行うことができる。記録媒体読取装置 164 は、記録媒体 167 に記録されているプログラムやデータを読み取るための装置である。読み取られたプログラムやデータはメモリ 162 や記憶装置 168 に格納される。従って、例えば記録媒体 167 に記録された管理プログラム 169 やプログラム 813 等を、記録媒体読取装置 164 を用いて上記記録媒体 167 から読み取って、メモリ 162 や記憶装置 168 に格納するようにすることができる。記録媒体 167 としてはフレキシブルディスクや CD-ROM、DVD-ROM、半導体メモリ等を用いることができる。記録媒体読取装置 164 は管理端末 160 に内蔵されている形態とすることもできるし、外付されている形態とすることもできる。記憶装置 168 には管理プログラム 169 が記憶されている。記憶装置 168 は、例えばハードディスク装置や半導体記憶装置等である。入力装置 165 はオペレータ等による管理端末 160 へのデータ入力等のために用いられる。入力装置 165 としては例えばキーボードやマウス等が用いられる。出力装置 166 は情報を外部に出力するための装置である。出力装置 166 としては例えばディスプレイやプリンタ等が用いられる。ポート 163 はディスク制御装置 110 と通信を行うための装置である。また図示されていない他のコンピュータとの間で通信を行うために使用することもできる。この場合、例えばプログラム 813 をポート 163 を介して他のコンピュータから受信して、ストレージ制御ユニット 800 にインストールするようにすることもできる。

【0077】

===コントローラの増設===

上述したように、本実施の形態に係るストレージシステム 100 においては、

ストレージ制御ユニット 8 0 0、チャネル制御ユニット 3 0 0、ディスク制御ユニット 4 0 0、グローバルキャッシュ 6 0 0 をディスク制御装置 1 1 0 に混在させて装着させることが可能である。これにより、顧客毎に異なるストレージシステム 1 0 0 の構成に対する要求に柔軟に対応することが可能となる。例えば、ストレージシステム 1 0 0 の導入時には少数のディスクドライブ 1 2 1 とストレージ制御ユニット 8 0 0 等により小規模のストレージシステム 1 0 0 を構成し、顧客の事業の拡大等の際にストレージ制御ユニット 8 0 0 や、チャネル制御ユニット 3 0 0、ディスク制御ユニット 4 0 0、グローバルキャッシュ 6 0 0 を追加するようにする。これにより顧客の要望に応じてストレージシステム 1 0 0 の規模拡大を行うことができる。その様子を図 1 7 に示す。またストレージシステム 1 0 0 の規模拡大前の導入時のシステム構成図を図 1 8 に示す。

【 0 0 7 8 】

図 1 8 に示す例では、導入時のディスク制御装置 1 1 0 は、クラスタを組むストレージ制御ユニット 8 0 0 と、内部接続部 5 0 0 と、管理端末 1 6 0 とにより構成されている。図 1 8 に示すディスク制御装置 1 1 0 において、導入時から高価な内部接続部 5 0 0 が用いられているのは、後から内部接続部 5 0 0 を追加保守するのは困難であるほか、ディスク制御装置 1 1 0 を大掛かりに分解した後の再組み立てが必要となり事実上不可能なためである。しかし図 1 9 に示すように導入時には内部接続部 5 0 0 を設けないように構成するようにすることも可能である。しかしながらこの場合、例えばグローバルキャッシュ 6 0 0 を増設する場合には、図 2 0 に示すように内部接続部 5 0 0 の増設も必要となる。

【 0 0 7 9 】

また、ストレージシステム 1 0 0 の導入をより容易に行うことができるように、初期コントローラ 1 1 1 を採用するようにすることもできる。その様子を図 2 1 乃至図 2 3 に示す。

【 0 0 8 0 】

図 2 1 に示すように、導入時には初期コントローラ 1 1 1 と少数のディスクドライブ 1 2 1 で運用を開始し、その後のシステム拡大時に装着部 1 3 0 にストレージ制御ユニット 8 0 0 やディスク制御ユニット 4 0 0 等を増設するようにする

こともできる。初期コントローラ 1 1 1 は、図 2 2 に示すように一つの筐体のなかにストレージ制御ユニット (S A) 8 0 0 や電源 1 1 2、ファン 1 1 3 が収納されたものであり、モジュラー型コントローラ 1 1 1 として提供されるものである。これにより、初期コントローラ 1 1 1 と必要な容量のディスクドライブ 1 2 1 を導入時に用意することにより、ストレージシステム 1 0 0 の運用を開始することができる。

【 0 0 8 1 】

図 2 1、図 2 2 に示す初期コントローラ 1 1 1 を用いた場合のストレージシステム 1 0 0 のシステム構成を図 2 3 に示す。図 2 3 に示すようにこの場合は、ストレージ制御ユニット 8 0 0 はモジュラー型コントローラ 1 1 1 すなわち初期コントローラ 1 1 1 として提供される。この場合ストレージシステム 1 0 0 の最大構成を想定したコストの高い内部接続部 5 0 0 を併せ持つ必要はない。すなわち一体型コントローラ 1 1 1 を導入する必要は無い。このため低コストでストレージシステム 1 0 0 を実現できる効果がある。またグローバルキャッシュ 6 0 0 やディスク制御ユニット 4 0 0 は装着部 1 3 0 に挿入される一体型コントローラとして提供される。この場合内部接続部 5 0 0 と初期コントローラ 1 1 1 間は、ケーブルにより接続されることになる。

【 0 0 8 2 】

また図 2 4 に示すように、ストレージシステム 1 0 0 の導入時に、初期コントローラ 1 1 1 の他に、モジュール化された S W (Switch) を装着するようにすることができる。S W は内部接続部 5 0 0 を構成する装置である。この場合のシステム構成図を図 2 5 に示す。図 2 5 に示すように、この場合の内部接続部 5 0 0 は、破線で囲んだように、モジュール化されて装着された S W と、一体型コントローラ 1 1 1 の導入時に追加された S W との組み合わせにより構成される。そして上記各 S W 間や S W とストレージ制御ユニット 8 0 0 間はケーブルにより接続されることになる。

【 0 0 8 3 】

===データ入出力処理の流れ===

次に、本実施の形態に係るストレージシステム 1 0 0 がホスト計算機 2 0 0 か

らデータ入出力要求を受信した場合に行われるデータ入出力処理の流れについて説明する。なお本実施の形態に係るデータ入出力処理は、各種の動作を行うためのコードから構成される制御プログラム 813、863、313、463を、それぞれプロセッサ 811、861、311、461が実行することにより実現される。

【0084】

まず本実施の形態に係るストレージシステム 100において、グローバルキャッシュ 600が増設された場合に行われる、ストレージ制御ユニット 800、チャンネル制御ユニット 300のローカルキャッシュ 830、330のボリウム管理テーブル 835、335のグローバルキャッシュ 600への移行処理の流れについて、図 26を参照しながら説明する。この処理は、管理端末 160からの指示により、ストレージ制御ユニット 800やチャンネル制御ユニット 300のプロセッサ 811、311により行われる。

【0085】

まずグローバルキャッシュ 600にボリウム管理テーブル 605を作成するための領域を確保する（S1000）。次にローカルキャッシュ 830、330のボリウム管理テーブル 835、335をロックする（S1001）。そしてローカルキャッシュ 830、330のボリウム管理テーブル 835、335の複製をグローバルキャッシュ 600に書き込む（S1002）。書き込みが終了したらローカルキャッシュ 830、330のボリウム管理テーブル 835、335のロックを解放して、処理を終了する（S1003）。なおS1000において確保する領域は、少なくともコピーするボリウム管理テーブル 835、335よりも大きくする。またロックする単位はコピー単位とすることもできる。

【0086】

これにより、例えばあるストレージ制御ユニット 800がホスト計算機 200からデータ入出力要求を受信した場合に、当該ストレージ制御ユニット 800のローカルキャッシュ 830を参照しても、データ入出力要求の対象となっている記憶ボリウムに対するユニットが特定できなかった場合に、グローバルキャッシュ 600のボリウム管理テーブル 605を参照することにより記憶ボリウムに対

する入出力制御を行うユニットを特定することができる。

【0087】

次に新規に記憶ボリウムが設定された場合に行われる、ボリウム管理テーブル 835、335 の更新処理について図 27 を参照しながら説明する。

まず、新規に設定された記憶ボリウムに対する入出力制御を行うユニットの種類が、ストレージ制御ユニット 800 であるかどうかの判定を行う (S2000)。ストレージ制御ユニット 800 である場合には、ローカルキャッシュ 830 のボリウム管理テーブル 835 をロックする (S2001)。そして新規に追加された記憶ボリウムに関する情報をボリウム管理テーブル 835 に書き込む (S2002)。そして書き込みが終了したらローカルキャッシュ 830 のボリウム管理テーブル 835 のロックを解放する (S2003)。続いてグローバルキャッシュ 600 のボリウム管理テーブル 605 に対してもローカルキャッシュ 830 のボリウム管理テーブル 835 と同様の書き込みを行って処理を終了する (S2004 乃至 S2006)。なお S2000 において、ストレージ制御ユニット 800 でない場合には、グローバルキャッシュ 600 のボリウム管理テーブル 605 に対してのみ処理を行う。

【0088】

次に、ホスト計算機 200 から本実施例に係るストレージシステム 100 に対してデータアクセス要求があった場合の処理の流れを示すフローチャートを図 28 に示す。

ホスト計算機 200 からストレージ制御ユニット 800 またはチャネル制御ユニット 300 にデータのアクセス要求があると (ステップ S3000)、ホスト I/F 制御部 810、310 のプロセッサ 811、311 はアクセス要求の解析を行う。解析によりアクセスの種類 (リード要求、ライト要求) やアクセスするデータのアドレス等を判別する。

【0089】

続いてプロセッサ 811、311 は、アクセスの種類に応じて、図 38 に示すコマンドをキャッシュ制御部 820、320 のキャッシュ制御部 I/F 回路 821、321 に送信する。リード要求の場合は図 38 D のコマンドが送信され、ライ

ト要求の場合は図 3 8 A のコマンドと図 3 8 B のコマンド（データ）が送信される。なお、図 3 8 に示すコマンドのフォーマットを図 3 7 に示す。図 3 7 に示すように図 3 8 に示すコマンドはヘッダー部とペイロード部を備えている。ヘッダー部は転送先アドレス、転送元アドレス、転送長、パケット種別を備えている。パケットの種別は、図 3 8 に示すように W R I T E 、 R E A D 、データ、ステータス情報である。

【 0 0 9 0 】

キャッシュ制御部 I F 回路 8 2 1 、 3 2 1 はプロセッサ 8 1 1 、 3 1 1 から送信されたコマンドに従い、ローカルキャッシュ 8 3 0 、 3 3 0 の制御領域 8 3 2 、 3 3 2 に記録されているボリウム管理テーブル 8 3 5 を参照して、データ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行うユニットを特定する。そしてボリウム管理テーブル 8 3 5 の ” A c c e s s M e t h o d ” 欄を参照して、当該ユニットに対するデータアクセス方法を特定する（ S 3 0 0 1 ）。「 d i r e c t 」と記載されていた場合には、 S 3 0 0 2 に進む。

【 0 0 9 1 】

S 3 0 0 2 において、キャッシュ制御部 I F 回路 8 2 1 、 3 2 1 はプロセッサ 8 1 1 、 3 1 1 から送信されたコマンドに従い、ローカルキャッシュ 8 3 0 、 3 3 0 の制御領域 8 2 2 、 3 3 2 に記録されているキャッシュデータ管理テーブル 8 3 4 、 3 3 4 を検索して、コマンドに指定されたアドレスのデータがローカルキャッシュ 8 3 0 、 3 3 0 に記憶されているかどうかを確認する（ステップ S 3 0 0 2 ）。

【 0 0 9 2 】

当該データがローカルキャッシュ 8 3 0 、 3 3 0 にある（ヒット）場合は（ S 3 0 0 2 ）、ローカルキャッシュ 8 3 0 、 3 3 0 に対するリードライト処理を行い（ S 3 0 0 7 ）、ホスト計算機 2 0 0 へ完了報告を行う（ S 3 0 0 8 ）。

【 0 0 9 3 】

S 3 0 0 2 において行われる処理を図 2 9 に示す。まず S 4 0 0 0 においてホスト計算機 2 0 0 からのデータ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行うユニットを判定する（ S 4 0 0 0 ）。自ユニット配下の記

憶ボリウムに対するものである場合には、自ボリウム用領域（第 1 の記憶領域）を対象にヒットミスの判定を行う（S 4 0 0 1）。ミスヒットの場合はグローバルキャッシュ 6 0 0 または記憶ボリウムから当該データをステージングする（S 4 0 0 2、S 4 0 0 3）。ステージングとは下位階層の記憶装置からデータを読み出してくることをいう。一方 S 4 0 0 0 において、他ユニット配下の記憶ボリウムに対するものである場合には、他ボリウム用領域（第 2 の記憶領域）を対象にヒットミスの判定を行う（S 4 0 0 4）。ミスヒットの場合はグローバルキャッシュ 6 0 0 または記憶ボリウムから当該データをステージングする（S 4 0 0 5、S 3 0 0 3）。

【 0 0 9 4 】

また S 3 0 0 7 において行われるローカルキャッシュ 8 3 0、3 3 0 に対するリードライト処理の流れについて図 3 3 を参照しながら説明する。

ホスト計算機 2 0 0 からのアクセス要求がリード要求の場合であれば、キャッシュ制御部 I F 回路 8 2 1、3 2 1 は当該データをローカルキャッシュ 8 3 0、3 3 0 から読み出してホスト計算機 2 0 0 へ送信する（S 8 0 0 0、S 8 0 0 1）。ローカルキャッシュ 8 3 0、3 3 0 から読み出し完了の報告（ACK）を受けると、キャッシュ制御部 I F 回路 8 2 1、3 2 1 はプロセッサ 8 1 1、3 1 1 に対してステータスを送信する。送信されるステータスは図 3 8 F で示されるコマンドである。最後にプロセッサ 8 1 1、3 1 1 はホスト計算機 2 0 0 にデータの読み出し完了報告を行って（S 3 0 0 8）処理を終了する。以上のリード要求の処理をフローチャートで表したものを図 4 0 に示す。

【 0 0 9 5 】

一方、ホスト計算機 2 0 0 からのアクセス要求がライト要求の場合は、キャッシュ制御部 I F 回路 8 2 1、3 2 1 は、ホスト計算機 2 0 0 から送信されバッファメモリ 8 2 2、3 2 2 に格納されている書き込みデータをローカルキャッシュ 8 3 0、3 3 0 に書き込む（S 8 0 0 2）。ローカルキャッシュ 8 3 0、3 3 0 への書き込み処理の詳細は図 3 4 に示される。すなわち、まずキャッシュ制御部 I F 回路 8 2 1、3 2 1 は、ペアとなっている相手側のキャッシュ制御部 I F 回路 8 2 1、3 2 1 に対してローカルキャッシュ 8 3 0、3 3 0 をロックするよう

に要求を出す。相手からロック確保の応答を受け取り、自他ローカルキャッシュ 3 3 0 のロックを確保すると (S 9 0 0 0)、キャッシュ制御部 I F 回路 8 2 1、3 2 1 は、バッファメモリ 8 2 2、3 2 2 に格納されている書き込みデータを、ペア間接続部 8 5 0、3 5 0 を介して相手側のバッファメモリ 8 2 2、3 2 2 に送信する。そして相手側のキャッシュ制御部 I F 回路 8 2 1、3 2 1 により相手側のローカルキャッシュ 8 3 0、3 3 0 に書き込みが行われる (S 9 0 0 1)。続いて自分側のローカルキャッシュ 8 3 0、3 3 0 にデータの書き込みを行う (S 9 0 0 2)。なお、データをローカルキャッシュ 8 3 0、3 3 0 に書き込む際にはキャッシュデータ管理テーブル 8 3 4、3 3 4 の” D i r t y ” 欄にチェックを入れる。相互のローカルキャッシュ 8 3 0、3 3 0 にデータの書き込みが完了すると、ロックを解除した後、ホスト計算機 2 0 0 へ完了報告を送信し、処理を終了する (S 9 0 0 3)。以上のライト要求の処理をフローチャートで表したものを図 4 0 に示す。

【 0 0 9 6 】

なお本実施例においては、ローカルキャッシュ 8 3 0、3 3 0 のキャッシュデータ管理テーブル 8 3 4、3 3 4 の検索や、ローカルキャッシュ 8 3 0、3 3 0 からのデータの読み出し等の制御は、キャッシュ制御部 I F 回路 8 2 1、3 2 1 が行う場合を例に説明したが、プロセッサ 8 1 1、3 1 1 が行う態様とすることも可能である。

【 0 0 9 7 】

また詳細は後述するが、グローバルキャッシュ 6 0 0 へのデータアクセス制御についても、キャッシュ制御部 I F 回路 8 2 1、3 2 1 が行う態様に限らず、プロセッサ 8 1 1、3 1 1 が行う態様とすることも可能である。

【 0 0 9 8 】

次に、ホスト計算機 2 0 0 からのデータ入出力要求を受けたが、ローカルキャッシュ 8 3 0、3 3 0 に当該データが無い場合、すなわちキャッシュミスヒットの場合の処理について説明する。

この場合はグローバルキャッシュ 6 0 0 に当該データがあるかどうかを確認する (S 3 0 0 3)。まずキャッシュ制御部 I F 回路 8 2 1、3 2 1 はプロセッサ

8 1 1、3 1 1 から送信されたコマンドに指定された当該データのアドレスを元に、内部接続部 5 0 0 を介してグローバルキャッシュ 6 0 0 にコマンドを送信する。そしてグローバルキャッシュ 6 0 0 の制御領域 6 0 2 に記録されているキャッシュデータ管理テーブル 6 0 4 検索して、当該データがグローバルキャッシュ 6 0 0 に記憶されているかどうかを確認する。

【 0 0 9 9 】

当該データがグローバルキャッシュ 6 0 0 に無ければ、ボリウム管理テーブル 6 0 5 を参照して、データ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行うユニットに対してコマンドを送信する。そして当該データを記憶ボリウムから読み出して、グローバルキャッシュ 6 0 0 に格納させる（S 3 0 0 4）。グローバルキャッシュ 6 0 0 に格納されたデータは、内部接続部 5 0 0 を介してもう 1 つのグローバルキャッシュ 6 0 0 へも送られ、データの 2 重化が行われる。

【 0 1 0 0 】

なおここで、記憶ボリウムからグローバルキャッシュ 6 0 0 へ読み出したデータをいち早くホスト計算機 2 0 0 へ届けるための処理を優先させ、グローバルキャッシュ 6 0 0 上でのデータの 2 重化を後回しにする態様とすることも可能である。かかるグローバルキャッシュ 6 0 0 上のデータは記憶ボリウムにも記憶されているので、グローバルキャッシュ 6 0 0 上で消失しても問題は生じないからである。当該データが更新された場合に 2 重化を行うようにすることでデータの信頼性は確保できる。

【 0 1 0 1 】

続いて当該データをグローバルキャッシュ 6 0 0 上でロックを掛ける（S 3 0 0 5）。すなわち、グローバルキャッシュ 6 0 0 上の当該データを他のローカルキャッシュ 8 3 0、3 3 0 から読み出されないようにする。その処理の流れを図 3 0 のフローチャートに示す。

当該データがすでに他のローカルキャッシュ 8 3 0、3 3 0 に読み出されており、ロックが掛けられている場合には（S 5 0 0 0）、当該ローカルキャッシュ 8 3 0、3 3 0 に対してロックを解放するように要求する（S 5 0 0 1）。どの

ローカルキャッシュ 8 3 0、3 3 0 がロックを掛けているのかは、キャッシュデータ管理テーブル 6 0 4 の” O w n e r ” 欄で知ることができる。ロックが解放されるのを待った後（S 5 0 0 2）、ロックを掛け、他のローカルキャッシュ 8 3 0、3 3 0 から読み出されないようにしてから処理を終了する（S 5 0 0 3）。他のローカルキャッシュ 8 3 0、3 3 0 にロックが掛けられていなければ、直ちにロックを掛けて処理を終了する（S 5 0 0 0、S 5 0 0 3）。

【 0 1 0 2 】

続いてグローバルキャッシュ 6 0 0 上に格納された当該データをローカルキャッシュ 8 3 0、3 3 0 へ読み出す処理を行う（S 3 0 0 6）。その処理の流れを図 3 1 のフローチャートに示す。

まずグローバルキャッシュ 6 0 0 からローカルキャッシュ 8 3 0、3 3 0 にデータを転送する前に、ローカルキャッシュ 8 3 0、3 3 0 上に当該データを書き込むためのキューの空きスロットがあるかどうかを調べる（S 6 0 0 0）。ここでスロットとはキューを構成する個々の記憶領域を言う。なおこの処理は、ローカルキャッシュ 8 3 0、3 3 0 上に当該データを書き込むための空き領域があるかどうかを調べるようにすることもできる。この場合は、キャッシュデータ管理テーブル 8 3 4、3 3 4 の” V a l i d ” 欄を検索して無効なデータの総容量がグローバルキャッシュ 6 0 0 から転送されるデータの総容量よりも大きいかどうかをチェックする。

【 0 1 0 3 】

空きスロットがある場合には、まずキャッシュ制御部 I F 回路 8 2 1、3 2 1 は、ペアとなっている相手側のキャッシュ制御部 I F 回路 8 2 1、3 2 1 に対してローカルキャッシュ 8 3 0、3 3 0 をロックするように要求を出してロックを確保する（S 6 0 0 2）。次にグローバルキャッシュ 6 0 0 からデータをバッファメモリ 8 2 2、3 2 2 に格納し、ペア間接続部 8 5 0、3 5 0 を介して相手側のバッファメモリ 8 2 2、3 2 2 にデータを送信するとともに、自分のローカルキャッシュ 8 3 0、3 3 0 にもデータの書き込みを行う（S 6 0 0 3、S 6 0 0 4）。相互のローカルキャッシュ 8 3 0、3 3 0 にデータの書き込みが完了すると、ロックを解除して処理を終了する（S 6 0 0 5）。この後の処理はホスト計

算機 2 0 0 からのデータ入出力要求に応じて、上述した通りに行われる（S 3 0 0 7、S 3 0 0 8）。

【0 1 0 4】

なお、グローバルキャッシュ 6 0 0 からローカルキャッシュ 8 3 0、3 3 0 にデータを転送するための空きスロットがない場合は、ローカルキャッシュ 8 3 0、3 3 0 上のいずれかのデータをグローバルキャッシュ 6 0 0 に書き戻すことにより空きスロットを確保する処理が必要になる（S 6 0 0 1）。その処理の流れを図 3 2 のフローチャートに示す。

まずキャッシュ制御部 I F 回路 8 2 1、3 2 1 は、ペアとなっている相手側のキャッシュ制御部 I F 回路 8 2 1、3 2 1 に対してローカルキャッシュ 8 3 0、3 3 0 をロックするように要求を出してロックを確保する（S 7 0 0 0）。続いて所定のアルゴリズムにより特定したグローバルキャッシュ 6 0 0 へ書き出されるデータの D i r t y ビットをキャッシュデータ管理テーブル 8 3 4、3 3 4 により調べる（S 7 0 0 1）。所定のアルゴリズムとしては、最も長期間アクセスのなかったデータをキャッシュから書き出す L R U（Least Recently Used）方式が一般的であるが、他のアルゴリズムとすることもできる。

【0 1 0 5】

D i r t y ビットがセットされていなければグローバルキャッシュ 6 0 0 へデータを書き出す必要はないが、D i r t y ビットがセットされている場合はグローバルキャッシュ 6 0 0 へデータを書き出す必要があるので、グローバルキャッシュ 6 0 0 に当該データを書き込むための空きスロットがあるかどうかを調べる（S 7 0 0 2）。グローバルキャッシュ 6 0 0 に空きスロットがない場合はグローバルキャッシュ 6 0 0 のデータを記憶ポリウムへ書き出して、空きスロットを確保する（S 7 0 0 3）。

【0 1 0 6】

続いてグローバルキャッシュ 6 0 0 上の空きスロットにローカルキャッシュ 8 3 0、3 3 0 からデータを書き出す（S 7 0 0 4）。書き出しは 2 つのグローバルキャッシュ 6 0 0 に対して行われる。グローバルキャッシュ 6 0 0 にデータが書き出された後は、もはや当該データは D i r t y ではないので、D i r t y ビ

ットをリセットする (S7005)。続いて当該データが記憶されていたローカルキャッシュ 830、330 上のスロットを解放する必要がある場合には (S7006)、当該データの Valid ビットをリセットする (S7007)。

【0107】

そしてペアとなっている相手のローカルキャッシュ 830、330 に対して当該データのグローバルキャッシュ 600 への書き出しが完了した旨の報告を行う (S7008)。この報告をうけた相手側のローカルキャッシュ 830、330 では、キャッシュデータ管理テーブル 834、334 の Valid ビットがリセットされる。最後にローカルキャッシュ 830、330 のロックを解放して (S7009) 処理を終了する。

【0108】

一方、S3001 においてボリウム管理テーブル 825、325 の "Access Method" 欄に "message" と記載されていた場合には、S3009 に進む。S3009 の処理の流れを図 35 に示す。

【0109】

まずキャッシュ制御部 IF 回路 821、321 は、メッセージアクセス用の領域 (通信バッファ) を確保する (S10000)。すなわち、データ入出力要求の対象となっている記憶ボリウムに対する入出力制御を行うユニットのローカルキャッシュ 830、330 の通信バッファ 837 の空き領域を確保する。

【0110】

ホスト計算機 200 からのデータ入出力要求がリード要求の場合には、上記領域を確保した通信バッファ 837 に当該データ入出力要求を書きこむ (S10001、S10003)。データ入出力要求の書きこみは、図 39 に示すメッセージにより行われる。図 39 A に示すメッセージにより、メッセージコマンドであることが示される。そして図 39 B に示すメッセージのメッセージデータ欄に当該データ入出力要求が挿入されて、上記通信バッファ 837 に書きこまれる。データ入出力制御の終了通知及び読み出されたデータを受け取ったら (S10004)、ホスト計算機 200 に当該データを送信する (S10006)。

【0111】

一方、ホスト計算機 2 0 0 からのデータ入出力要求がライト要求の場合には、上記領域を確保した通信バッファ 8 3 7 に当該データ入出力要求及び書き込みデータを書きこむ（S 1 0 0 0 1 乃至 S 1 0 0 0 3）。そして通信バッファ 8 3 7 にデータ入出力制御の終了通知が書き込まれたら（S 1 0 0 0 4）、ホスト計算機 2 0 0 に当該終了通知を送信して処理を終了する。

【 0 1 1 2 】

自己の通信バッファ 8 3 7 にメッセージを書き込まれたストレージ制御ユニット 8 0 0 により行われる処理を図 3 6 に示す。

まず、キャッシュ制御部 I F 回路 8 2 1、3 2 1 は、自己の通信バッファ 8 3 7 にメッセージを書き込まれたことを検知すると（S 1 1 0 0 1）、通信バッファ 8 3 7 からデータ入出力要求を読み出す（S 1 1 0 0 2）。次に当該データ入出力要求の対象となっているデータがローカルキャッシュ 8 3 0 に記憶されているか否かをチェックする（S 1 1 0 0 3）。ミスヒットの場合は、グローバルキャッシュ 6 0 0 または記憶ボリウムから当該データを読み出してきてローカルキャッシュ 8 3 0 に記憶する（S 1 1 0 0 4）。そしてデータ入出力要求がリード要求の場合は、当該読み出したデータをメッセージ送信元の通信バッファ 8 3 7 に書き込む（S 1 1 0 0 6）。またデータ入出力要求がライト要求の場合は、当該データ入出力要求に従ってデータをローカルキャッシュ 8 3 0 へ書き込む（S 1 1 0 0 7）。この処理は図 3 4 に示した処理と同様である。そして書き込み完了通知を相手側の通信バッファ 8 3 7 に書き込む。

【 0 1 1 3 】

なお、通信バッファ 8 3 7 を介してストレージ制御ユニット 8 0 0 間でメッセージの授受を行う際の処理の流れを図 4 1 に示す。

通信バッファ 8 3 7 を介してデータ入出力要求の授受を行うようにすることにより、各ストレージ制御ユニット 8 0 0 は、他のストレージ制御ユニット 8 0 0 の処理に依存せずに、当該他のストレージ制御ユニットに接続されている記憶ボリウムのデータに対するデータ入出力制御を行うことが可能となる。

【 0 1 1 4 】

次に、ストレージシステム 1 0 0 の導入時にストレージ制御ユニット 8 0 0 に

より運用を行っていたストレージシステム 100 において、ディスク制御ユニット 400 が増設された場合に行うことができる、記憶ボリュームの変更処理について図 42 を参照しながら説明する。

【0115】

この処理は、ストレージ制御ユニット 800 に接続されている記憶ボリュームに記憶されているデータの複製を、ディスク制御ユニット 400 に接続されている記憶ボリュームに書き込むことにより、それ以降のホスト計算機 200 からのデータ入出力要求に対する入出力制御を当該ディスク制御ユニット 400 に接続された記憶ボリュームに対して行うようにするものである。このようにすることにより、ストレージシステム 100 の構成変更の柔軟性を高めることができるようになる。例えば、ストレージシステム 100 の導入時には初期コントローラ 111 により運用していた顧客が、その後のストレージシステム 100 の規模拡大時に、チャンネル制御ユニット 300 とディスク制御ユニット 400 とを用いた、拡張性の高いシステム構成に変更するようにすることが出来るようになる。

【0116】

この処理は、管理端末 160 からの指示により、ストレージ制御ユニット 800 やチャンネル制御ユニット 300 のプロセッサ 811 により行われる。

まずプロセッサ 811 は、ローカルキャッシュ 830 のボリューム管理テーブル 835、及びグローバルキャッシュ 600 のボリューム管理テーブル 605 をロックする (S16000、S16001)。そして、ローカルキャッシュ 830 のボリューム管理テーブル 835 の "DA No" 欄、"volume No" 欄、"drive No" 欄を、ストレージ制御ユニット 800 に接続されている記憶ボリュームに関する情報から、ディスク制御ユニット 400 に接続されている記憶ボリュームに関する情報に変更する (S16002)。またグローバルキャッシュ 600 のボリューム管理テーブル 605 の "DA No" 欄、"volume No" 欄、"drive No" 欄を、ストレージ制御ユニット 800 に接続されている記憶ボリュームに関する情報から、ディスク制御ユニット 400 に接続されている記憶ボリュームに関する情報に変更する (S16003)。そしてローカルキャッシュ 830 のボリューム管理テーブル 835、及びグローバルキャッシュ

600のボリウム管理テーブル605をロックを解放する（S16004、S16005）。

【0117】

これにより、ホスト計算機200からのデータ入出力要求に対する入出力制御を、ストレージ制御ユニット800に接続された記憶ボリウムから、ディスク制御ユニット400に接続された記憶ボリウムに対して行うようにすることができるようになる。

【0118】

以上のように本実施の形態に係るディスク制御装置110においては、ストレージ制御ユニット800や、チャネル制御ユニット300、ディスク制御ユニット400、グローバルキャッシュ600のいずれも装着することができるので、顧客ニーズに応じた柔軟性の高いストレージシステム100構成することができる。これは、各ユニットのサイズやコネクタの位置、コネクタのピン配列等に互換性をもたせるようにしているからである。またボリウム管理テーブル835を設けることにより、異種ユニットが混在して装着されているディスク制御装置110においても、ホスト計算機200から送信されるデータ入出力要求に対して、各ユニットにおいてデータ入出力制御を行うことができるからである。

【0119】

また本実施の形態に係るディスク制御装置110においては、キャッシュ領域管理テーブル833を設けることにより、ストレージシステム100の性能向上を図ることができる。すなわちキャッシュ領域管理テーブル833の内容を変更することにより、ホスト計算機200から受信するデータ入出力要求の特性に適合した、キャッシュ領域を確保することが可能となる。これにより例えば、ホスト計算機200から受信したデータ入出力要求の多くが、当該データ入出力要求を受信したストレージ制御ユニット800に接続された記憶ボリウムに対するものである場合には、自SAVOL用領域836Aの割り当てを増やすようにすることにより、ホスト計算機200からのデータ入出力要求に対するキャッシュヒット率を増加させることができ、ストレージシステム100の性能向上を図ることが可能となる。

【0 1 2 0】

また本実施の形態に係るディスク制御装置 1 1 0 においては、グローバルキャッシュ 6 0 0 にもボリウム管理テーブル 6 0 5 を設けることにより、例えばあるストレージ制御ユニット 8 0 0 がホスト計算機 2 0 0 からデータ入出力要求を受信した場合に、当該ストレージ制御ユニット 8 0 0 のローカルキャッシュ 8 3 0 を参照しても、データ入出力要求の対象となっている記憶ボリウムに対するユニットが特定できなかった場合であっても、グローバルキャッシュ 6 0 0 のボリウム管理テーブル 6 0 5 を参照することにより記憶ボリウムに対する入出力制御を行うユニットを特定することができるようになる。

【0 1 2 1】

また他のストレージ制御ユニット 8 0 0 に接続された記憶ボリウムに対するデータ入出力処理を行う場合には、ストレージ制御ユニット 8 0 0 間で通信バッファ 8 3 7 を介してデータ入出力要求の授受を行うようにすることにより、各ストレージ制御ユニット 8 0 0 は、他のストレージ制御ユニット 8 0 0 の処理に依存せずに、当該他のストレージ制御ユニットに接続されている記憶ボリウムのデータに対するデータ入出力制御を行うことが可能となる。

【0 1 2 2】

また、ストレージ制御ユニット 8 0 0 に接続されている記憶ボリウムに記憶されているデータの複製を、ディスク制御ユニット 4 0 0 に接続されている記憶ボリウムに書き込むことにより、それ以降のホスト計算機 2 0 0 からのデータ入出力要求に対する入出力制御を当該ディスク制御ユニット 4 0 0 に接続された記憶ボリウムに対して行うようにすることができる。このようにすることにより、ストレージシステム 1 0 0 の構成変更の柔軟性を高めることができるようになる。例えば、ストレージシステム 1 0 0 の導入時には初期コントローラ 1 1 1 により運用していた顧客が、その後のストレージシステム 1 0 0 の規模拡大時に、チャネル制御ユニット 3 0 0 とディスク制御ユニット 4 0 0 とを用いた、拡張性の高いシステム構成に変更するようになることができるようになる。

【0 1 2 3】

以上本実施の形態について説明したが、上記実施例は本発明の理解を容易にす

るためのものであり、本発明を限定して解釈するためのものではない。本発明はその趣旨を逸脱することなく変更、改良され得ると共に、本発明にはその等価物も含まれる。

【 0 1 2 4 】

【発明の効果】

記憶デバイス制御装置、及び記憶デバイス制御装置の制御方法を提供することができる。

【図面の簡単な説明】

【図 1】 本実施の形態に係るストレージシステムの外観構成を示す図である。

【図 2】 本実施の形態に係るストレージシステムの全体構成を示すブロック図である。

【図 3】 本実施の形態に係るストレージ制御ユニットを示す図である。

【図 4】 本実施の形態に係るチャネル制御ユニットを示す図である。

【図 5】 本実施の形態に係るディスク制御ユニットを示す図である。

【図 6】 本実施の形態に係るキャッシュメモリユニットを示す図である。

【図 7】 本実施の形態に係るストレージシステムに、ストレージ制御ユニット、チャネル制御ユニット、ディスク制御ユニット、又はキャッシュメモリユニットが装着される様子を示す図である。

【図 8】 本実施の形態に係るストレージ制御ユニットの機能を示すブロック図である。

【図 9】 本実施の形態に係るストレージ制御ユニット間を接続するペア間接続部を説明するためのブロック図である。

【図 1 0】 本実施の形態に係るチャネル制御ユニットの機能を示すブロック図である。

【図 1 1】 本実施の形態に係るチャネル制御ユニット間を接続するペア間接続部を説明するためのブロック図である。

【図 1 2】 本実施の形態に係るディスク制御ユニットの機能を示すブロック図である。

【図 1 3】 本実施の形態に係るローカルキャッシュメモリの構成を示すブロック図である。

【図 1 4】 本実施の形態に係るグローバルキャッシュメモリの構成を示すブロック図である。

【図 1 5】 本実施の形態に係る内部接続部の構成を示すブロック図である。

【図 1 6】 本実施の形態に係る管理端末の構成を示すブロック図である。

【図 1 7】 本実施の形態に係るストレージシステムの規模を拡大する際の外観構成の変化の一例を示す図である。

【図 1 8】 本実施の形態に係るストレージシステムの規模を拡大する前の構成の一例を示す図である。

【図 1 9】 本実施の形態に係るストレージシステムの規模を拡大する前の構成の一例を示す図である。

【図 2 0】 本実施の形態に係るストレージシステムの規模を拡大した後の構成の一例を示す図である。

【図 2 1】 本実施の形態に係るストレージシステムの規模を拡大する際の外観構成の変化の一例を示す図である。

【図 2 2】 本実施の形態に係るストレージシステムにおける初期コントローラを示す図である。

【図 2 3】 本実施の形態に係るストレージシステムの規模を拡大する際の構成の一例を示す図である。

【図 2 4】 本実施の形態に係るストレージシステムの規模を拡大する際の外観構成の変化の一例を示す図である。

【図 2 5】 本実施の形態に係るストレージシステムの規模を拡大する際の構成の一例を示す図である。

【図 2 6】 本実施の形態に係るストレージシステムにおいて、ポリウム管理テーブルをローカルキャッシュメモリからグローバルキャッシュメモリへ移行する処理を示すフローチャートである。

【図 2 7】 本実施の形態に係るストレージシステムにおいて、ポリウムを

作成した場合に行われるボリウム管理テーブルを更新する処理を示すフローチャートである。

【図 2 8】 本実施の形態に係るデータアクセス処理を示すフローチャートである。

【図 2 9】 本実施の形態に係るローカルキャッシュメモリのヒットミス判定処理を示すフローチャートである。

【図 3 0】 本実施の形態に係るグローバルキャッシュメモリのロック確保処理を示すフローチャートである。

【図 3 1】 本実施の形態に係るローカルキャッシュメモリのステージング処理を示すフローチャートである。

【図 3 2】 本実施の形態に係るローカルキャッシュメモリのデステージング処理を示すフローチャートである。

【図 3 3】 本実施の形態に係るリードライト処理を示すフローチャートである。

【図 3 4】 本実施の形態に係るローカルキャッシュメモリへのライト処理を示すフローチャートである。

【図 3 5】 本実施の形態に係るメッセージ通信を送信する側によるデータアクセス処理を示すフローチャートである。

【図 3 6】 本実施の形態に係るメッセージ通信を受信する側によるデータアクセス処理を示すフローチャートである。

【図 3 7】 本実施の形態に係るコマンドの構成を示す図である。

【図 3 8】 本実施の形態に係るコマンドを示す図である。

【図 3 9】 本実施の形態に係るメッセージを示す図である。

【図 4 0】 本実施の形態に係るコマンドの送受信を示す図である。

【図 4 1】 本実施の形態に係るメッセージの送受信を示す図である。

【図 4 2】 本実施の形態に係るアクセス方法を変更する場合の処理を示す図である。

【符号の説明】

1 0 0 ストレージシステム

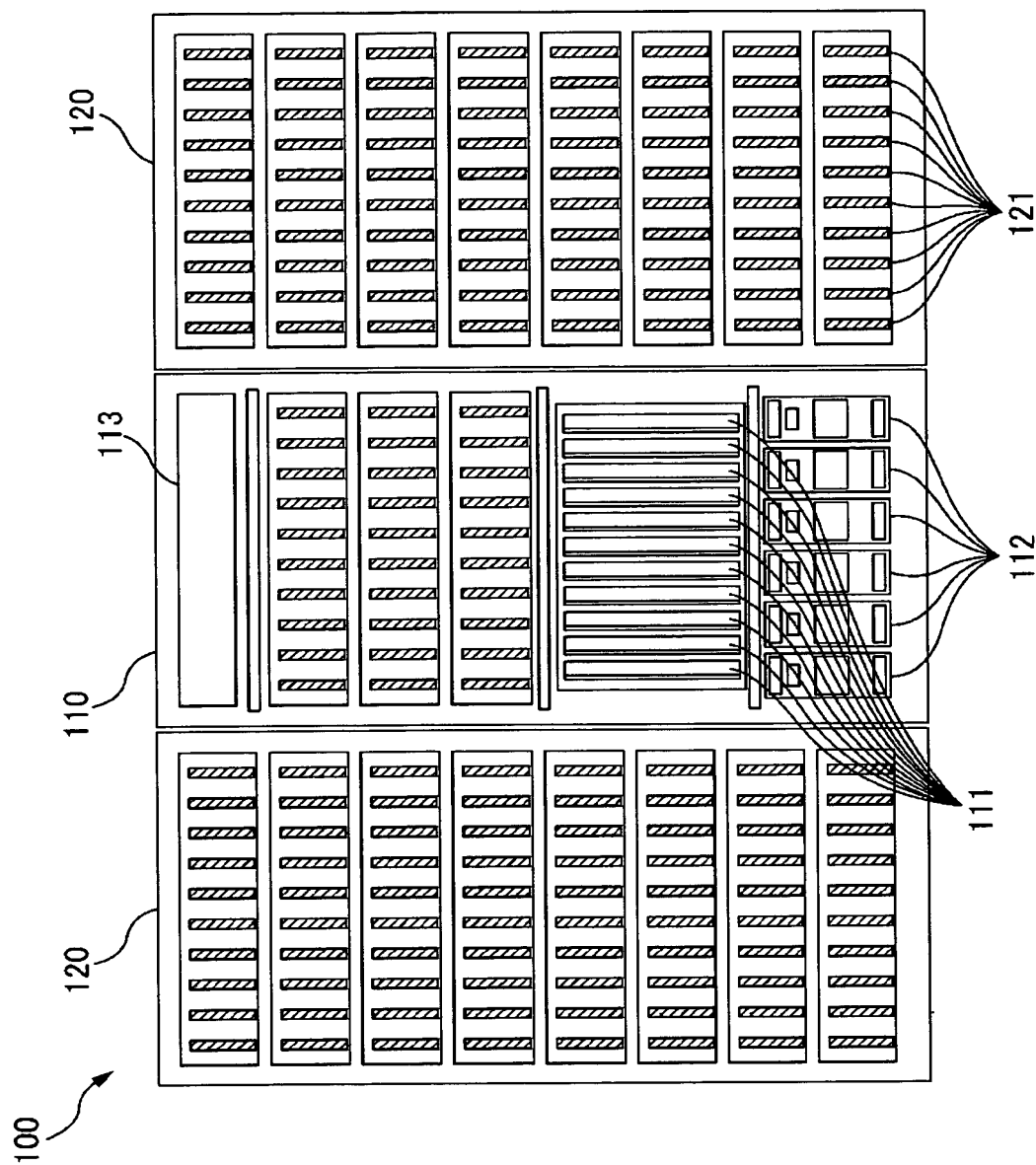
1 1 0 ディスク制御装置

1 2 0	ディスク駆動装置	1 3 0	装着部
1 6 0	管理端末	2 0 0	ホスト計算機
3 0 0	チャンネル制御ユニット	3 1 0	ホスト I F 制御部
3 3 0	ローカルキャッシュ	3 3 1	データ領域
3 3 2	制御領域	4 0 0	ディスク制御ユニット
4 6 0	ディスク I F 制御部	5 0 0	内部接続部
6 0 0	グローバルキャッシュ	6 0 3	キャッシュ領域管理テーブル
6 0 4	キャッシュデータ管理テーブル	6 0 5	ボリウム管理テーブル
6 0 6	ダイレクトアクセス用データ領域	6 0 7	通信バッファ
8 0 0	ストレージ制御ユニット	8 1 0	ホスト I F 制御部
8 3 0	ローカルキャッシュ	8 3 3	キャッシュ領域管理テーブル
8 3 4	キャッシュデータ管理テーブル	8 3 5	ボリウム管理テーブル
8 3 6	ダイレクトアクセス用データ領域		
8 3 7	通信バッファ	8 6 0	ディスク I F 制御部

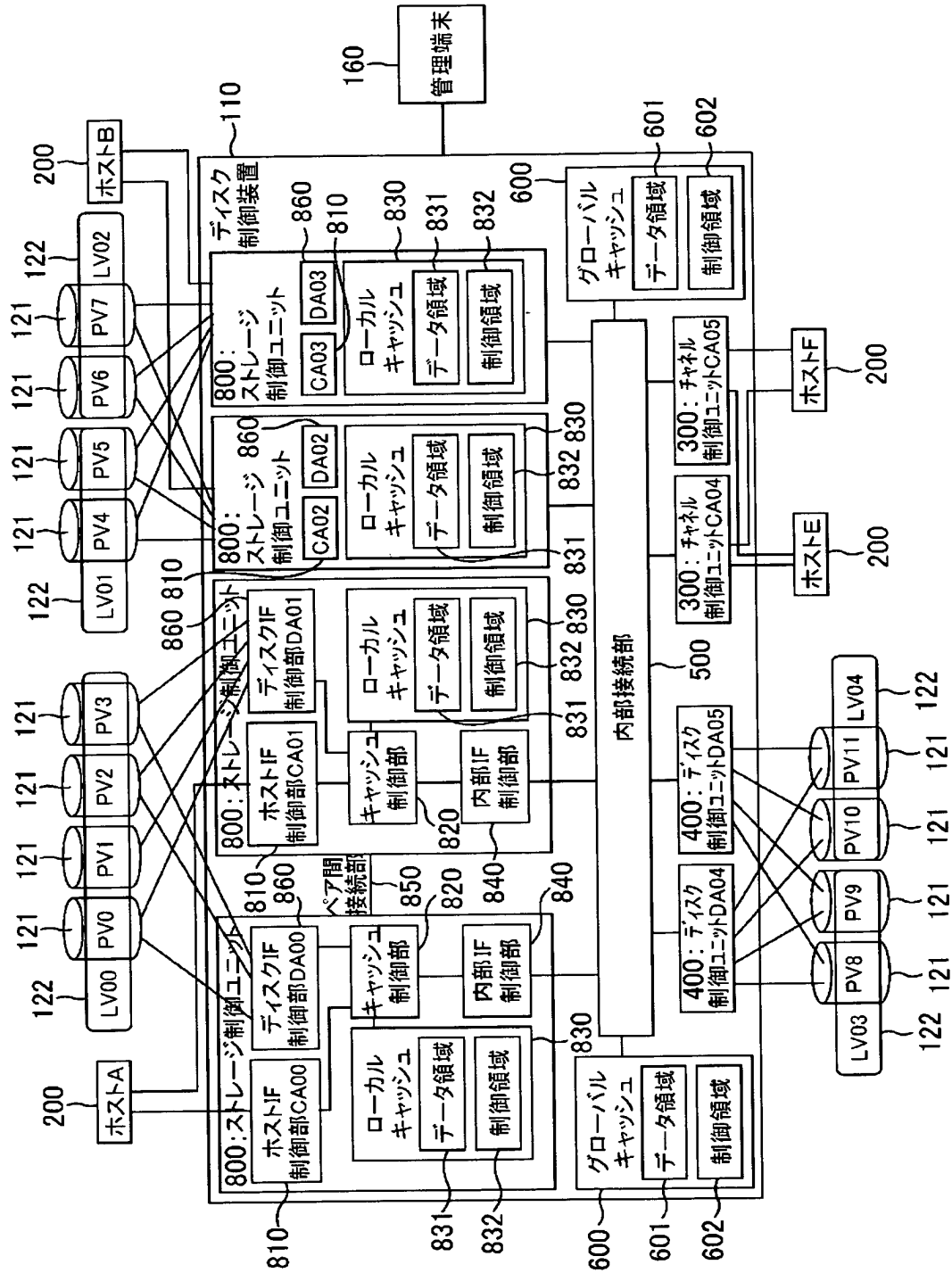
【書類名】

図面

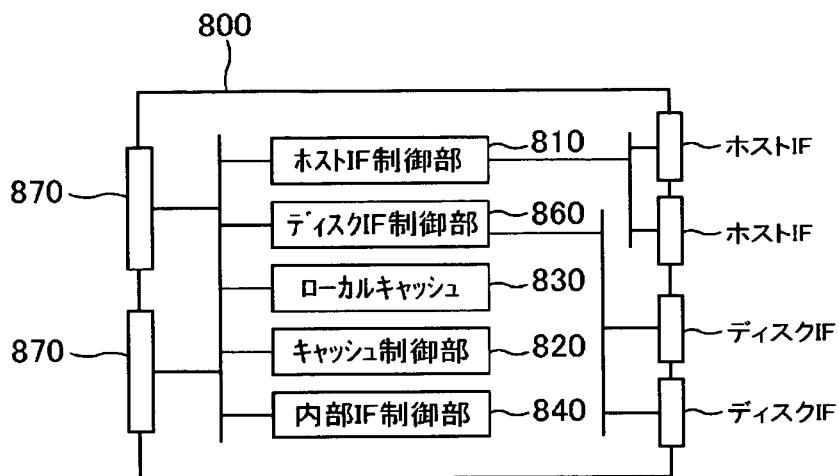
【図 1】



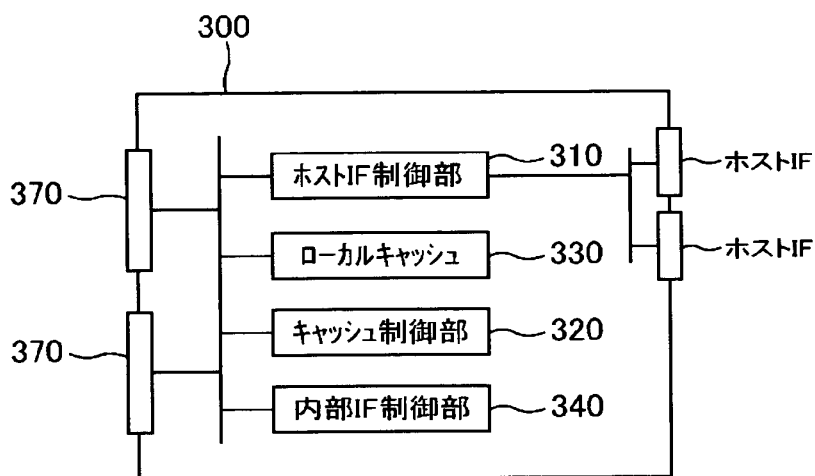
【図2】



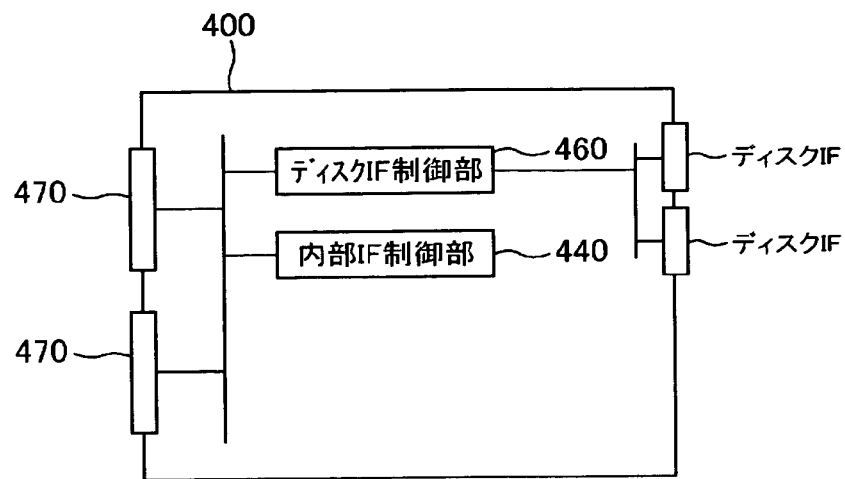
【図 3】



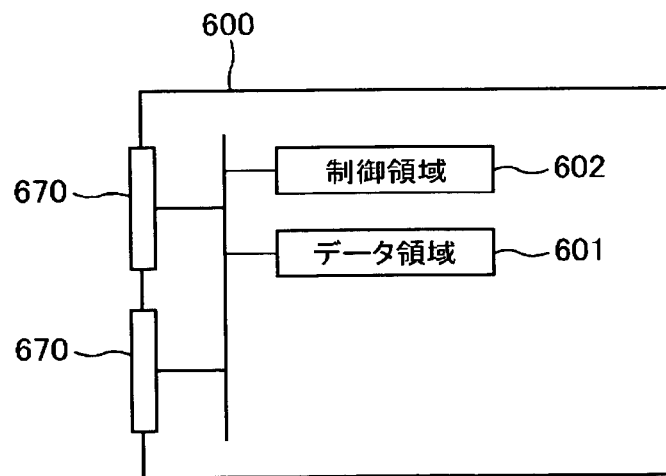
【図 4】



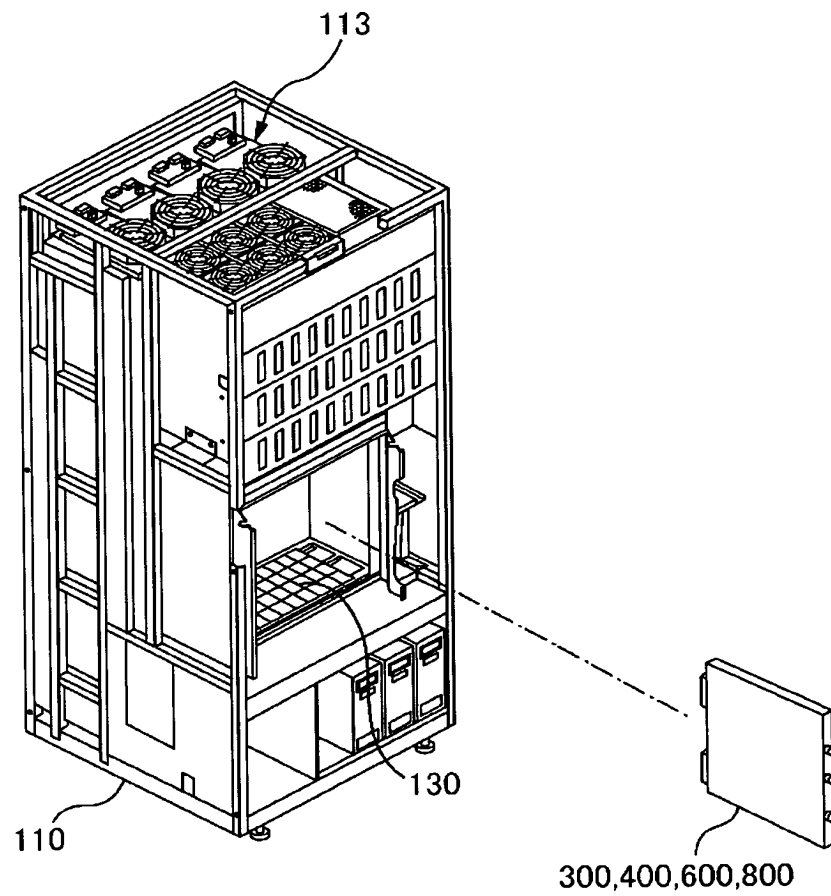
【図 5】



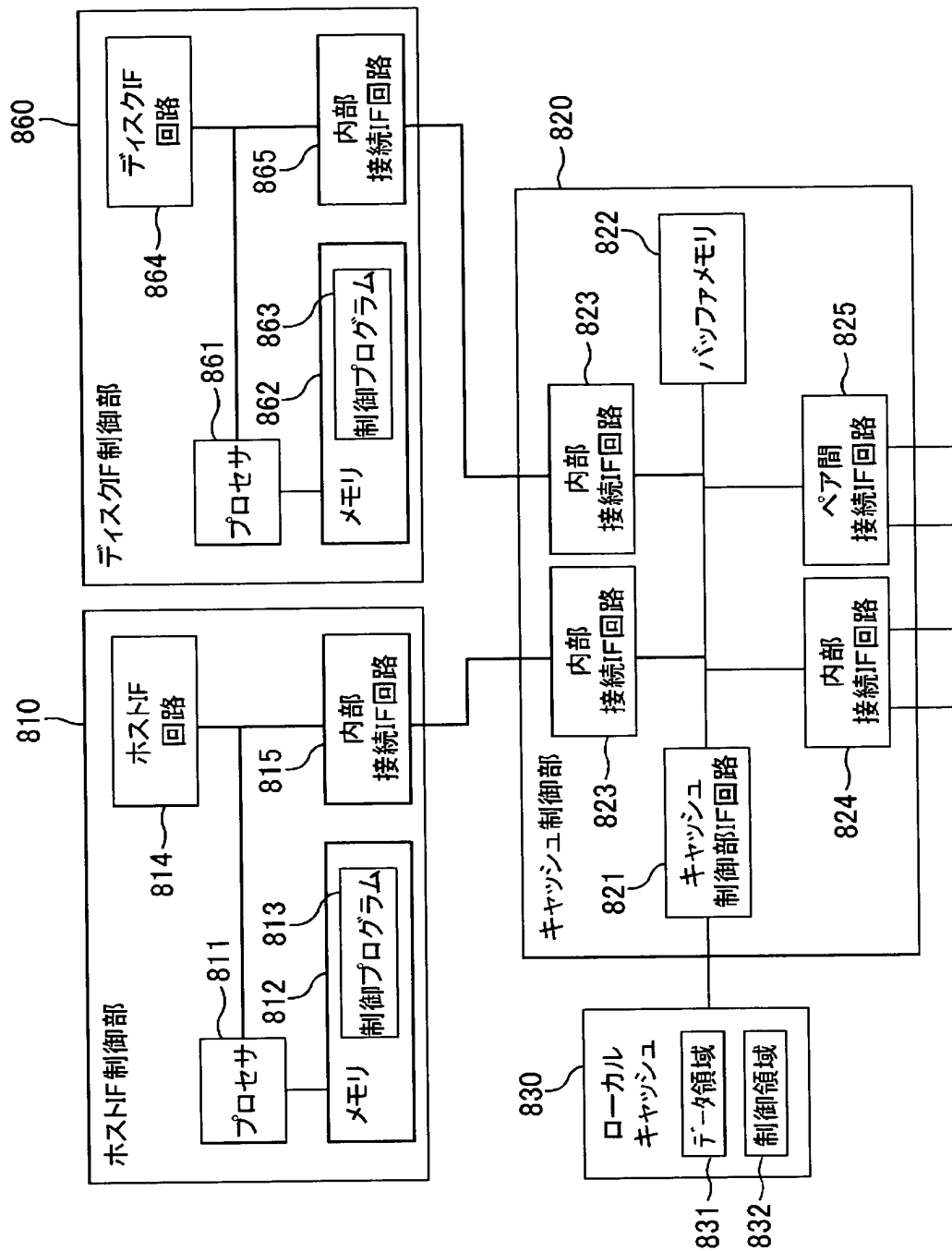
【図 6】



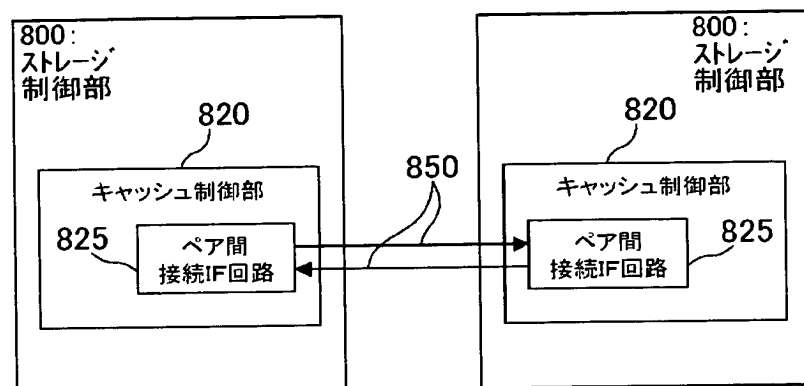
【図 7】



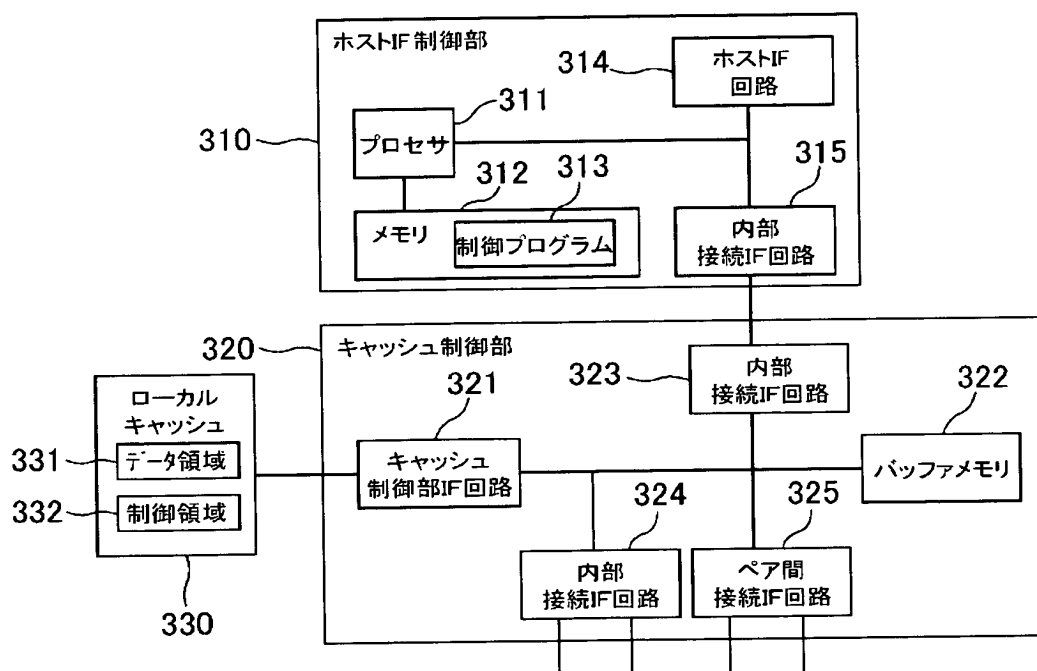
【図 8】



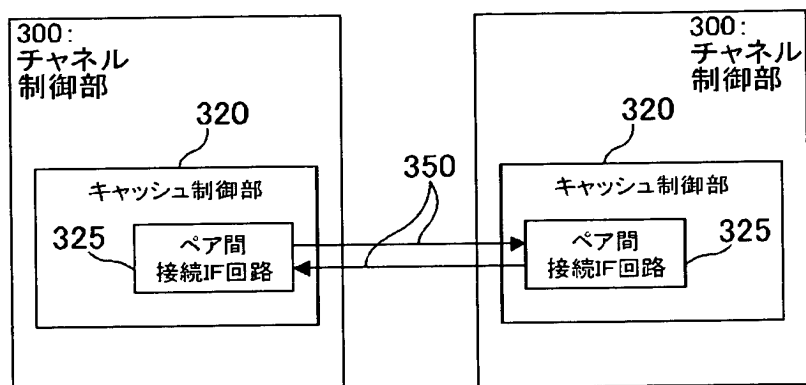
【図 9】



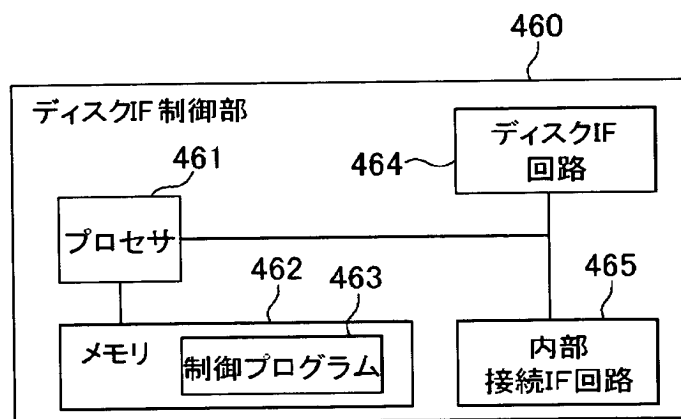
【図 10】



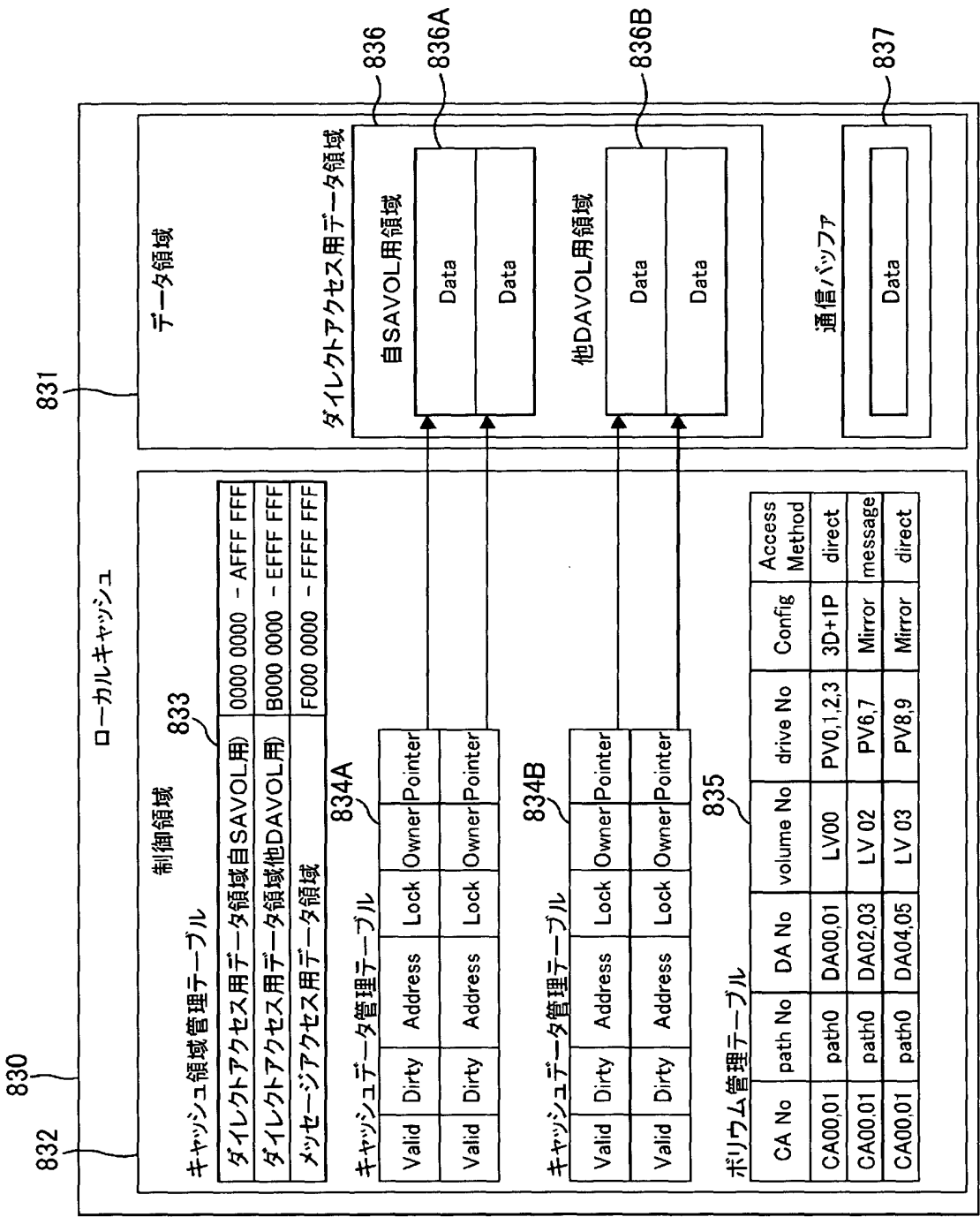
【図 1 1】



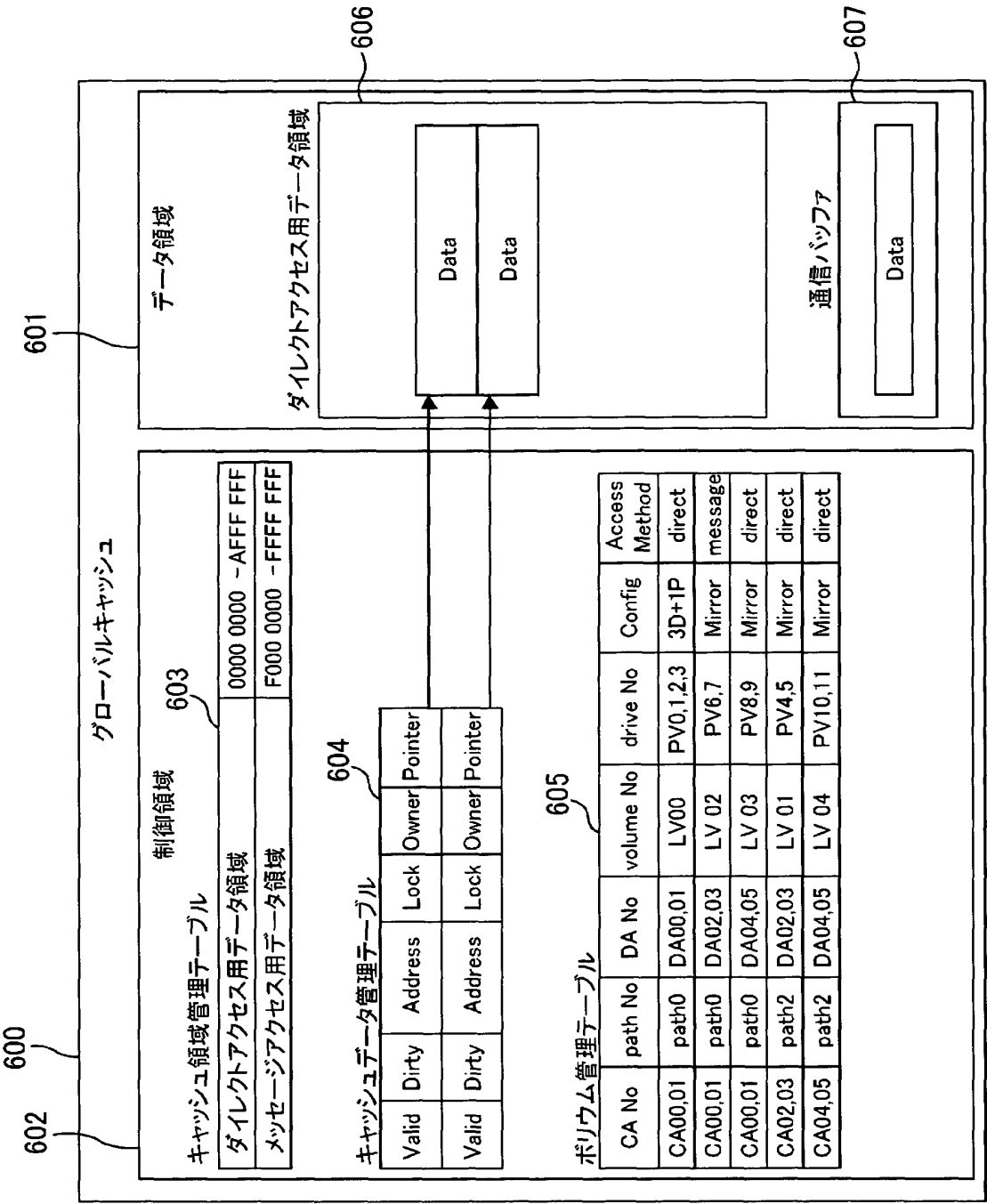
【図 1 2】



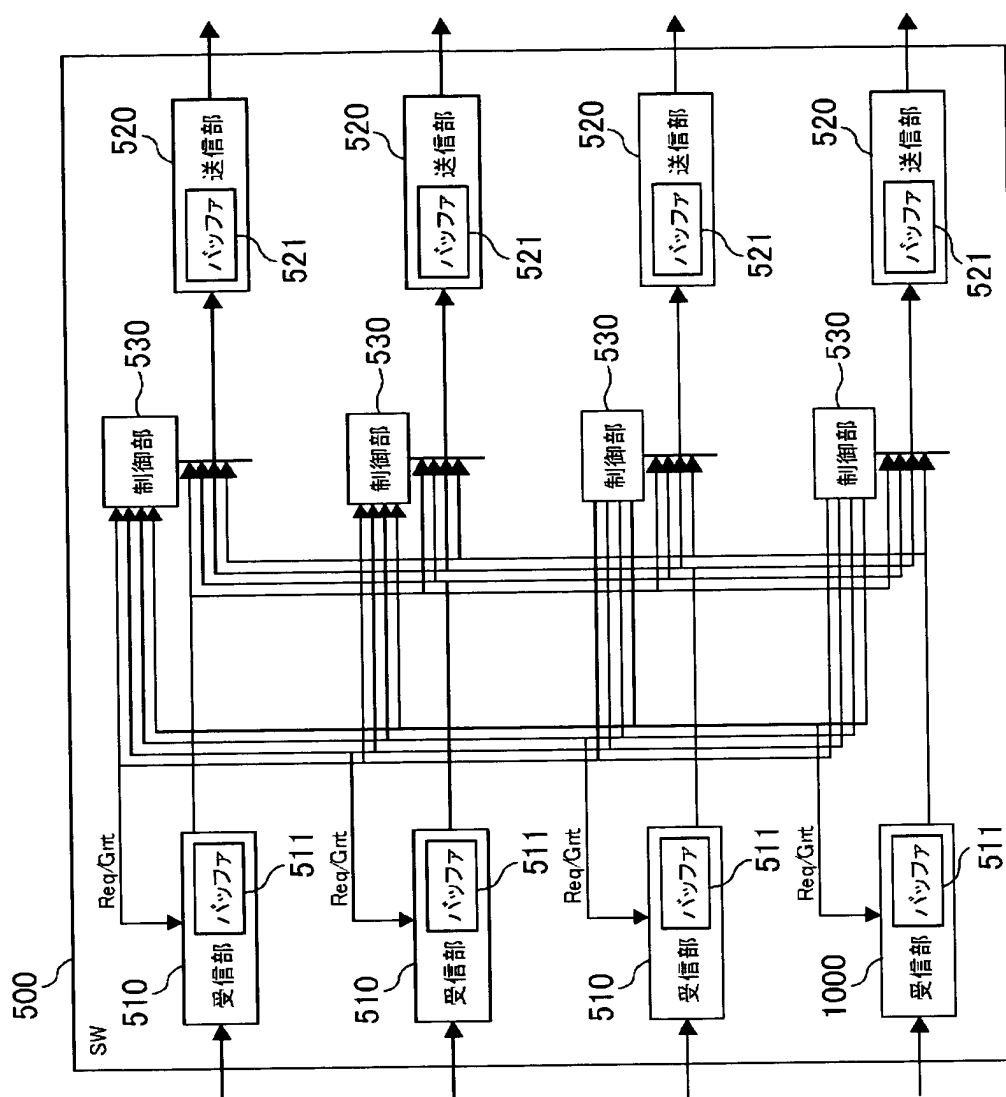
【図 13】



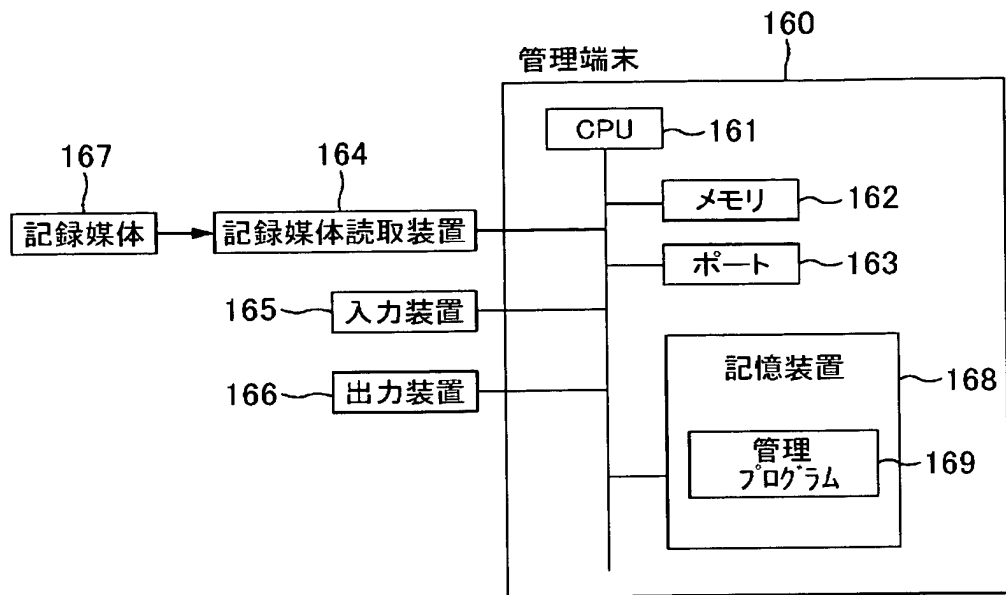
【図 14】



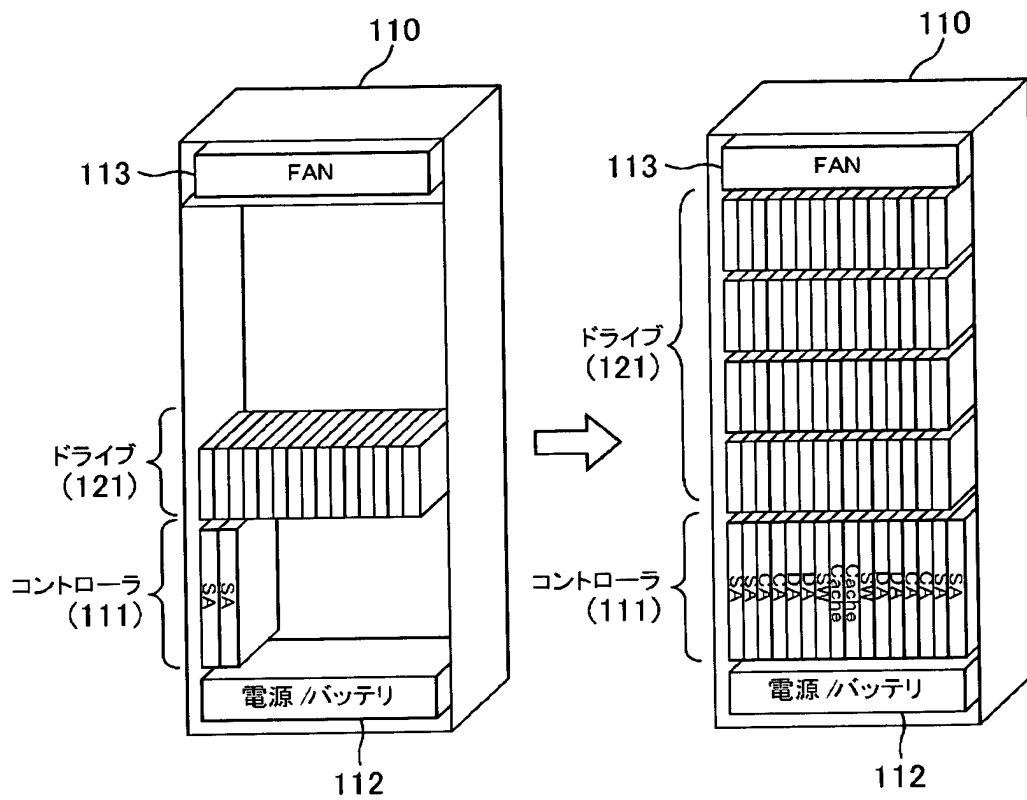
【図 15】



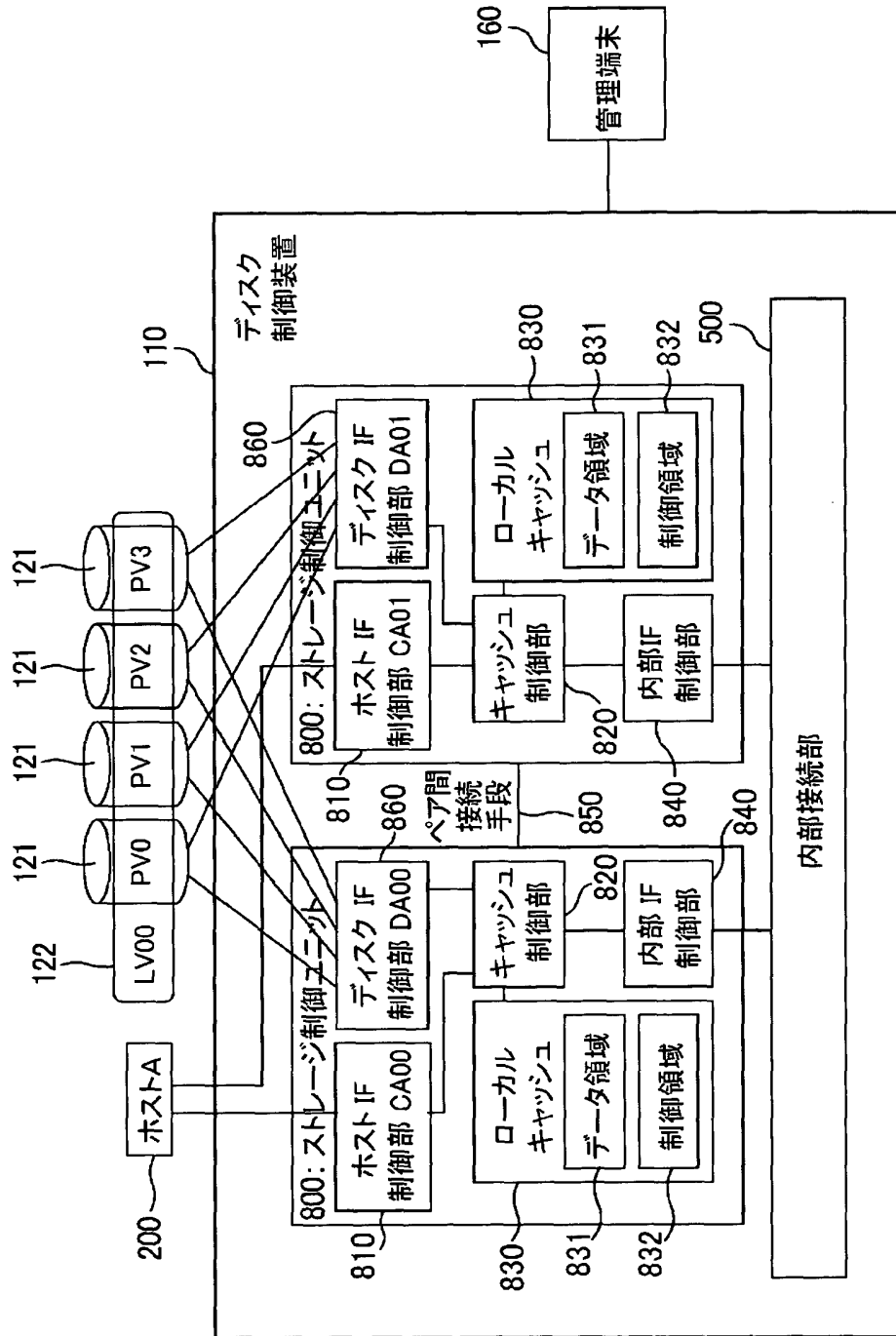
【図 16】



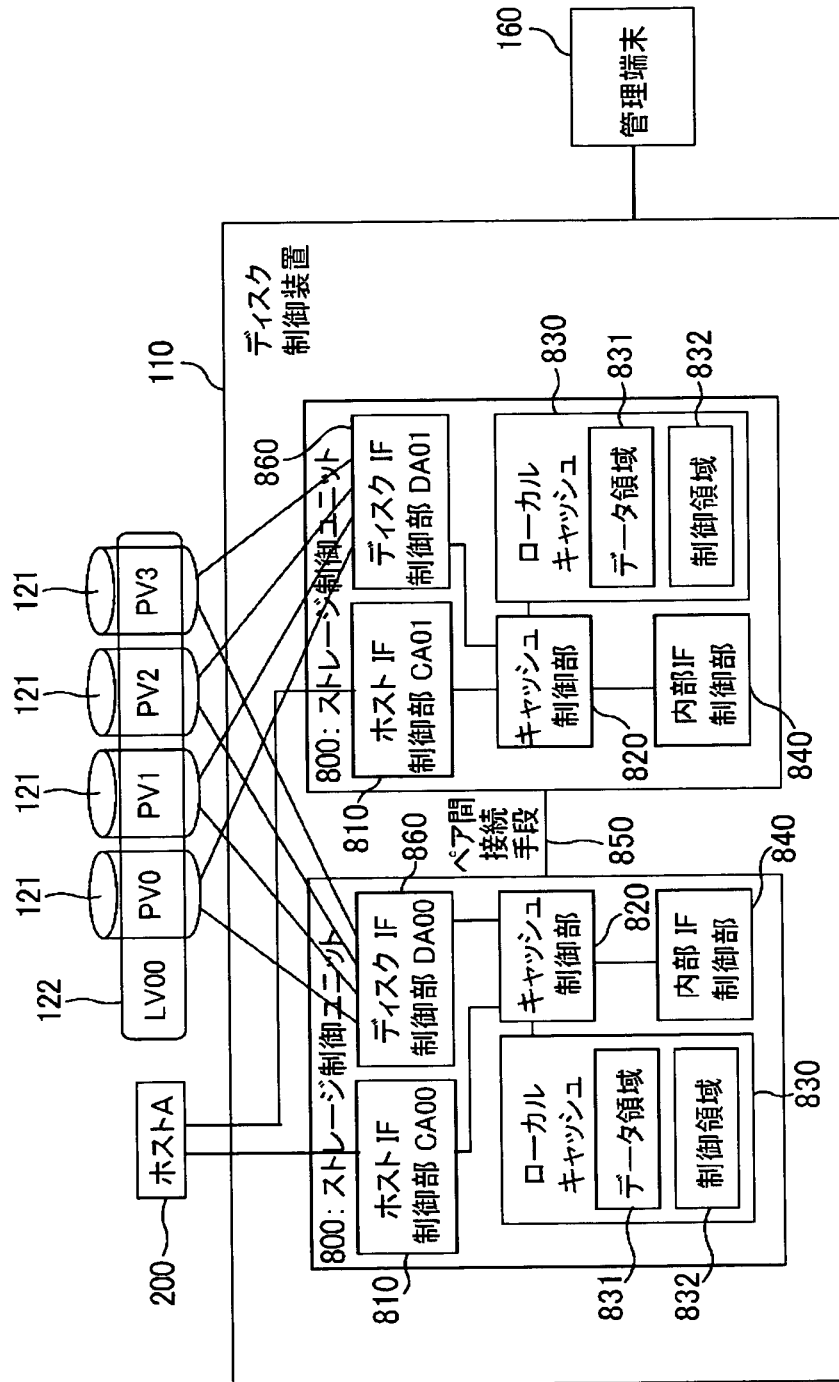
【図 17】



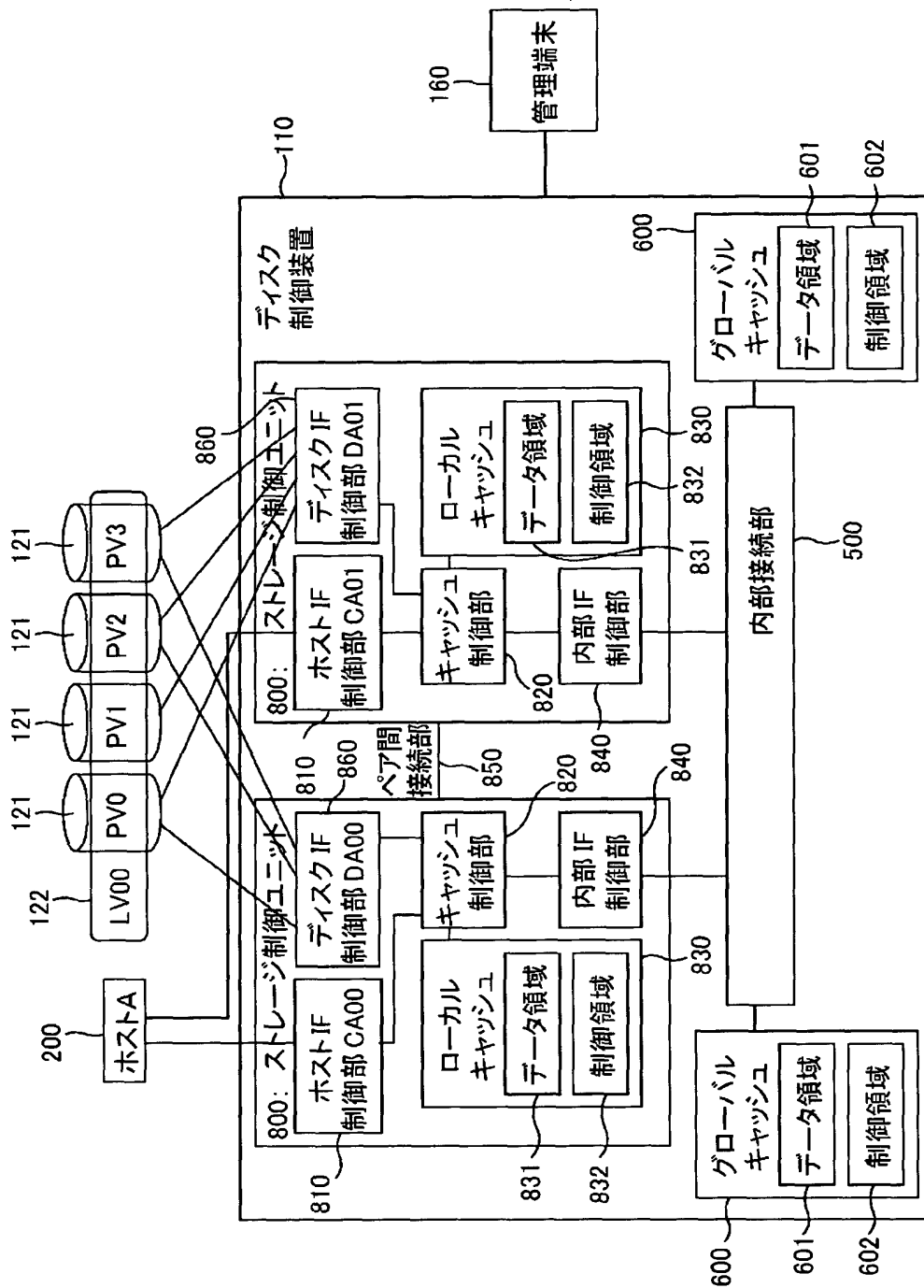
【図 18】



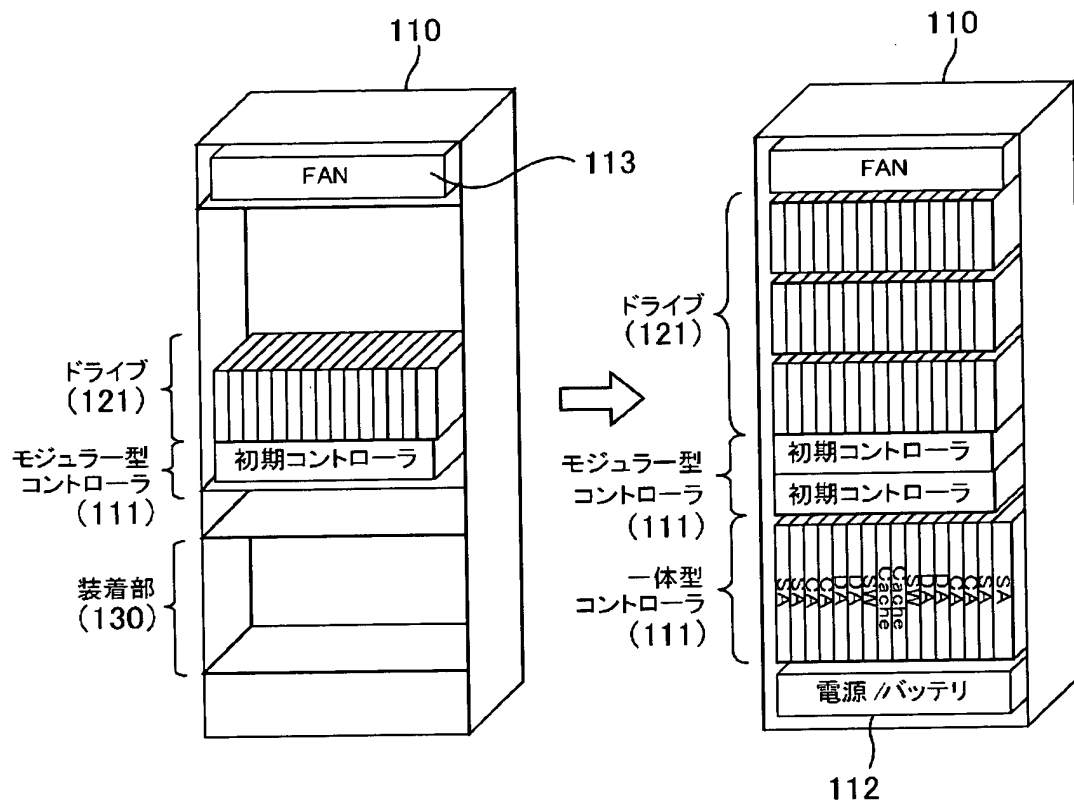
【図 19】



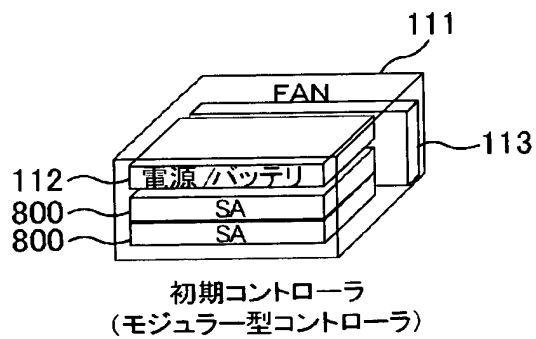
【図 20】



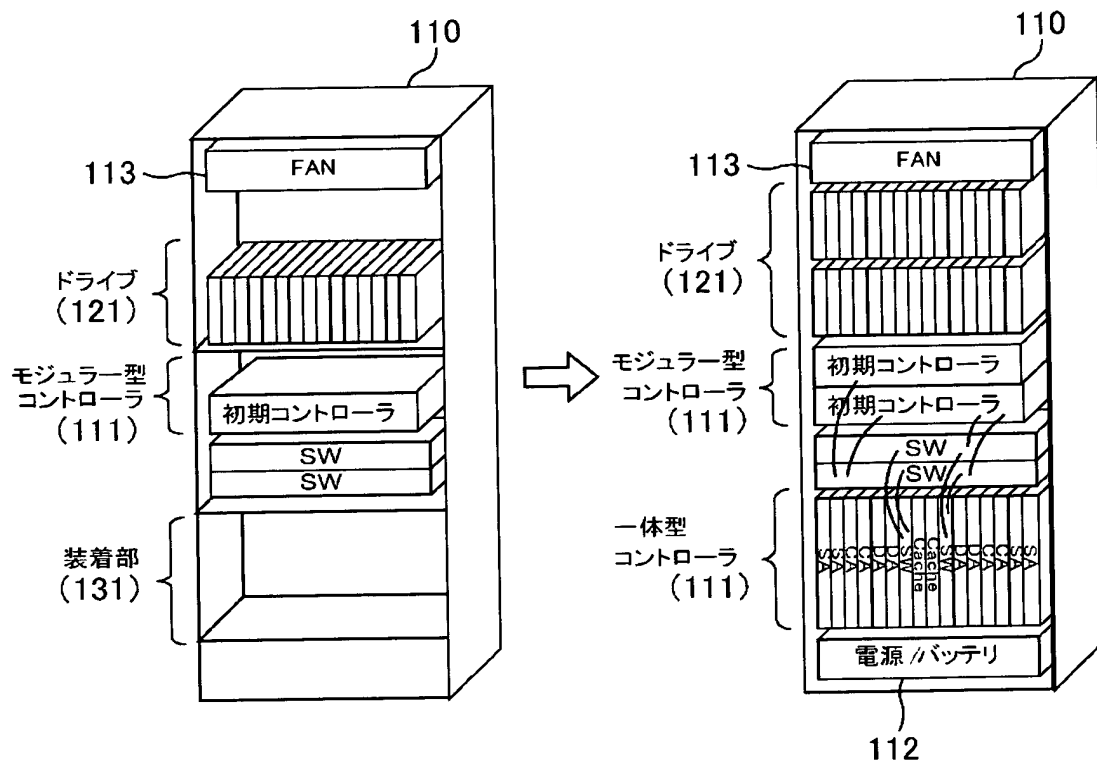
【図 2 1】



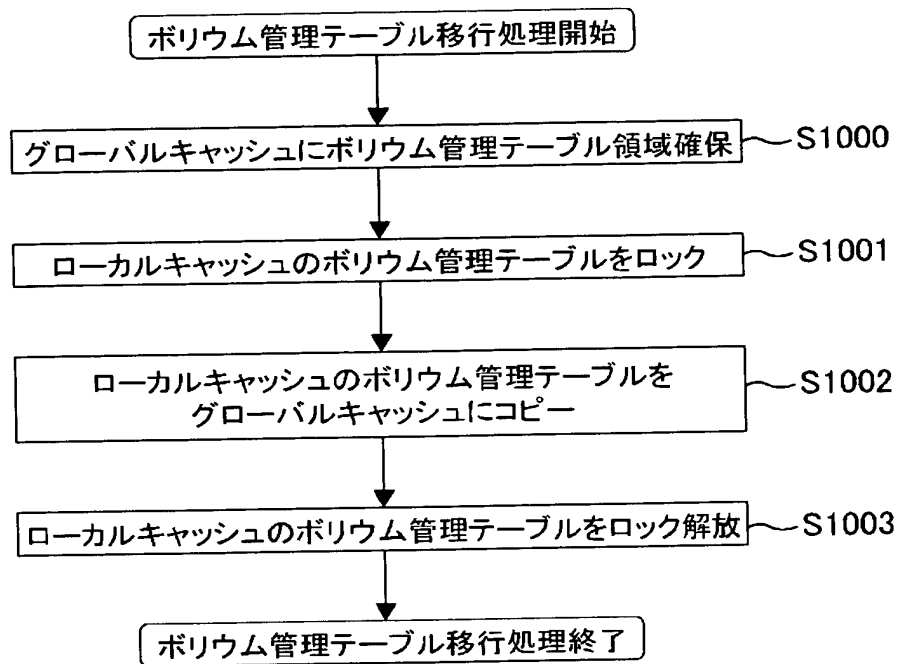
【図 2 2】



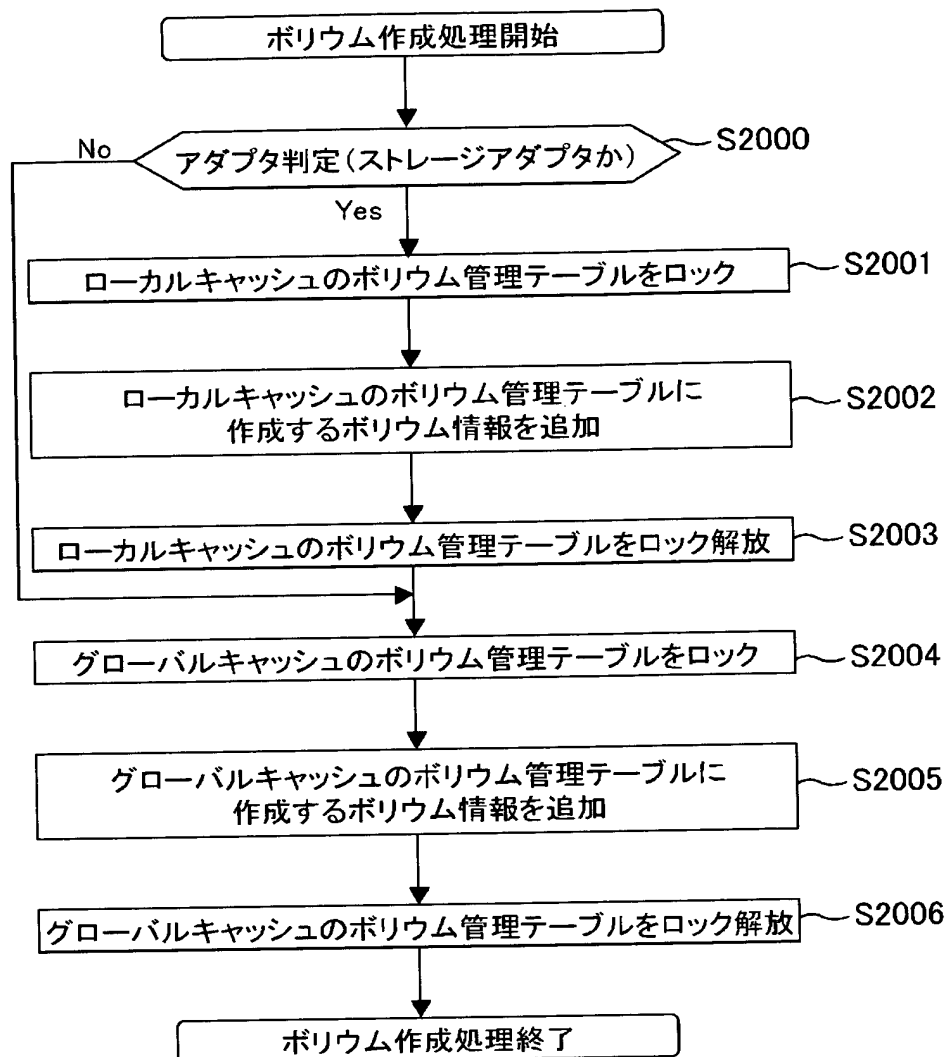
【図 24】



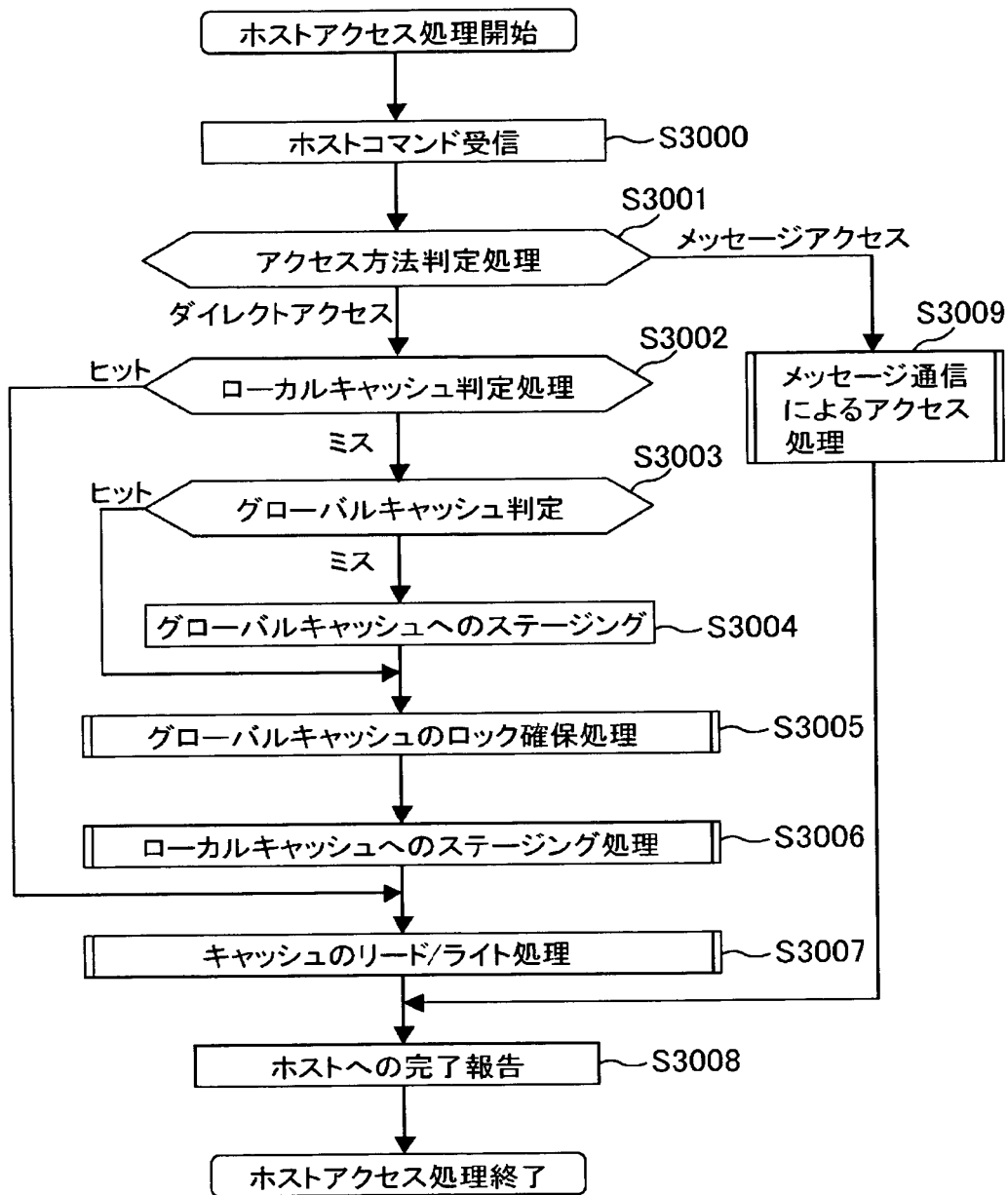
【図 26】



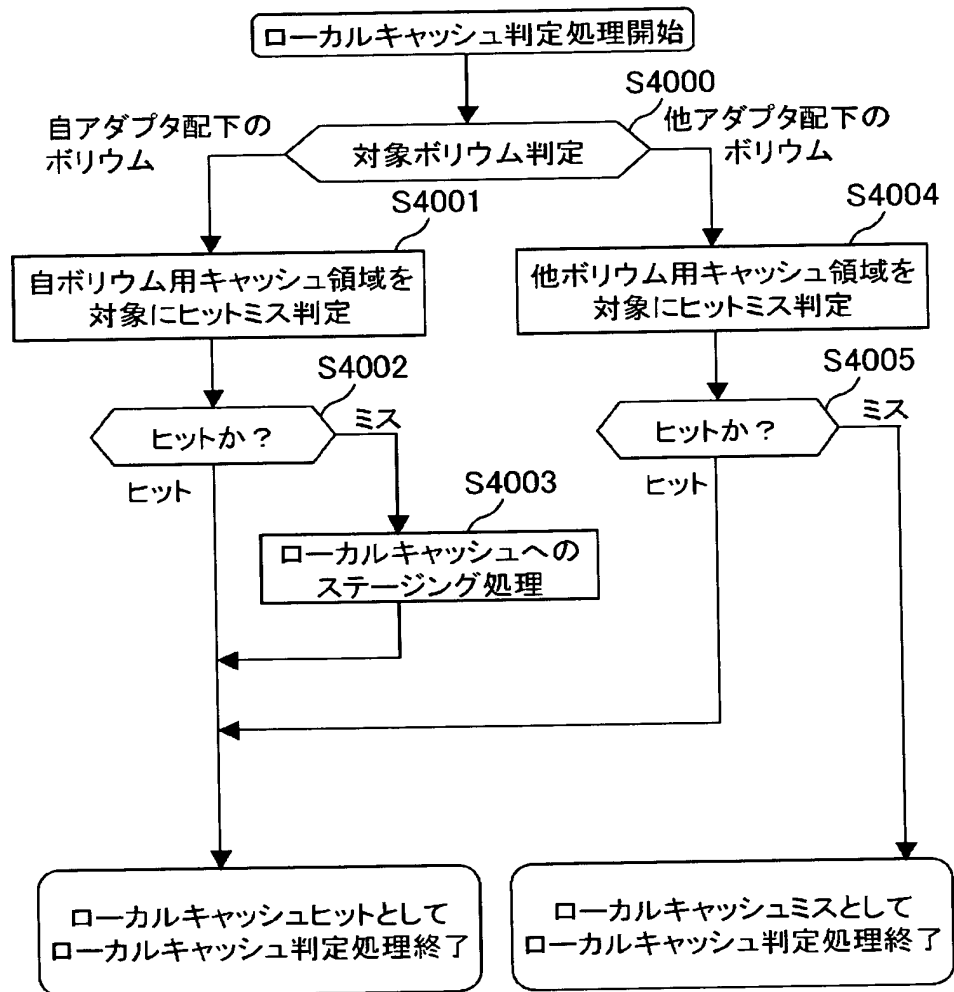
【図 27】



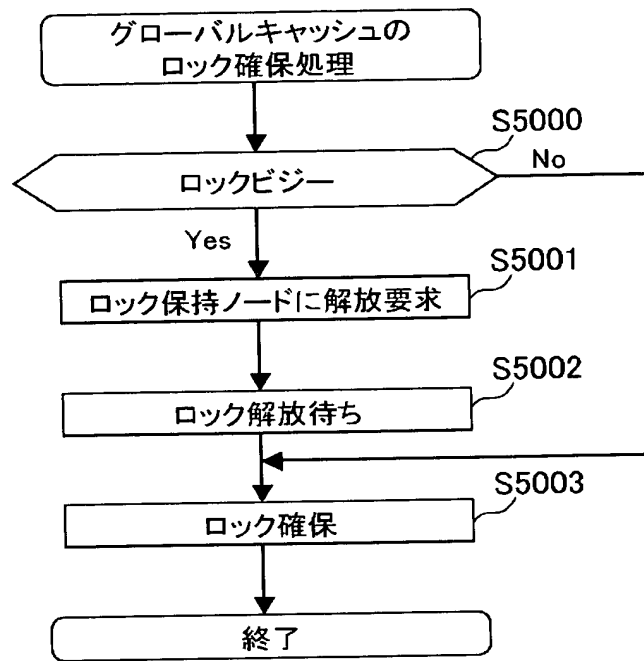
【図 28】



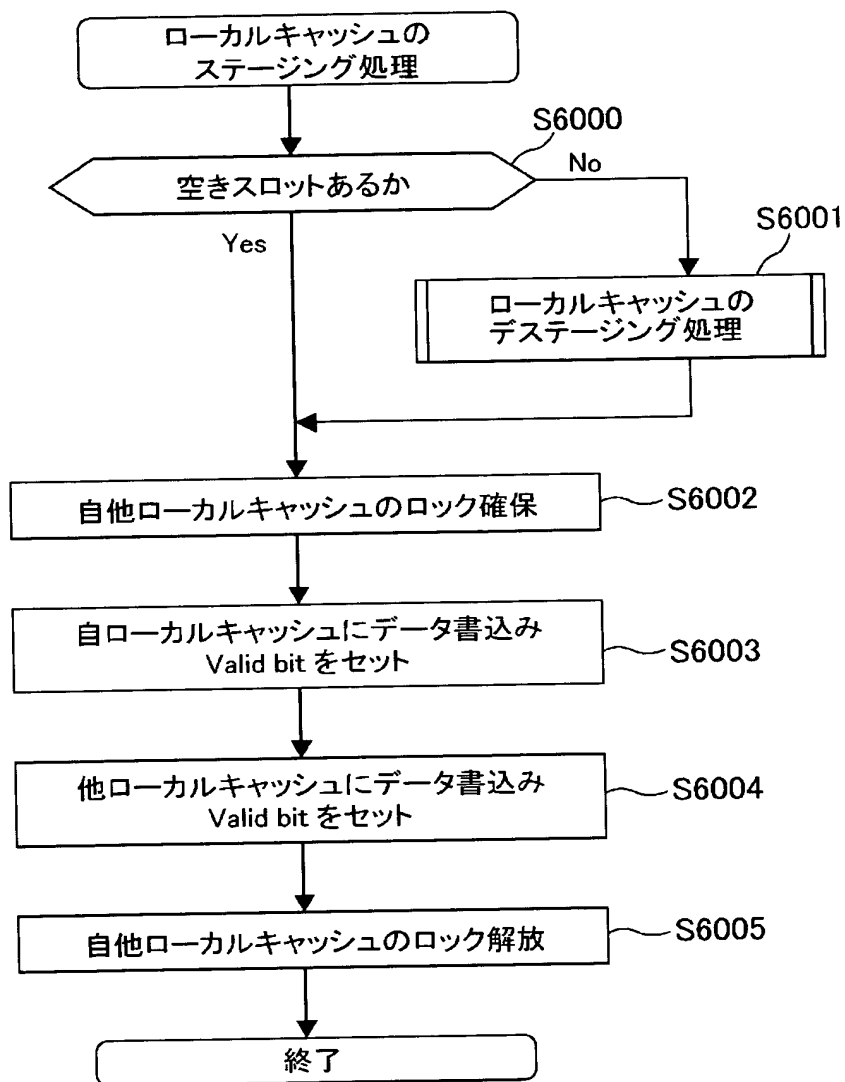
【図 29】



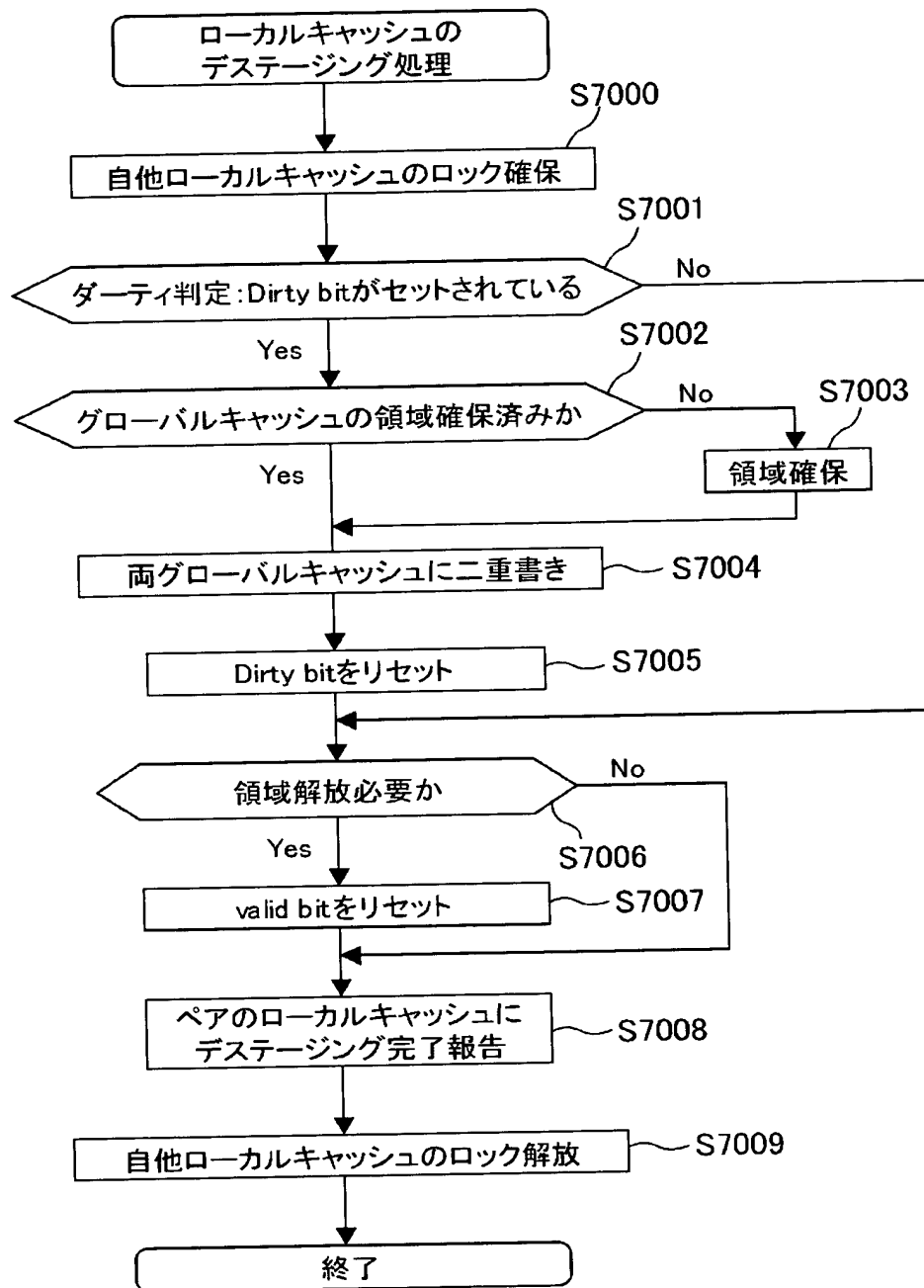
【図 30】



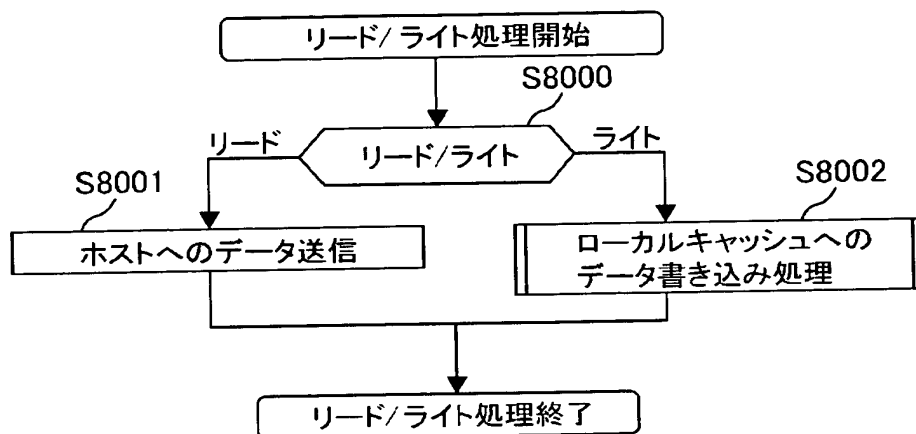
【図 31】



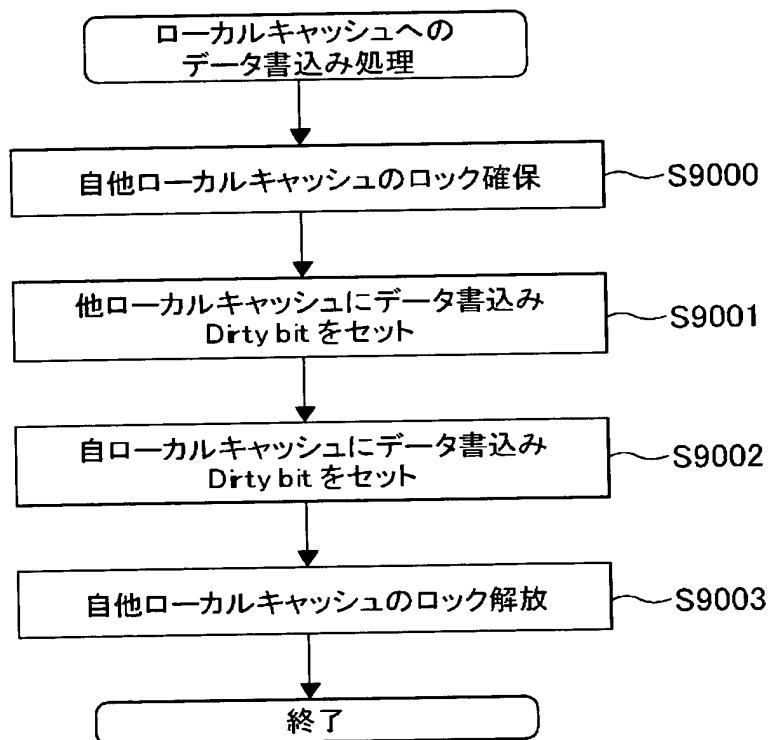
【図 3 2】



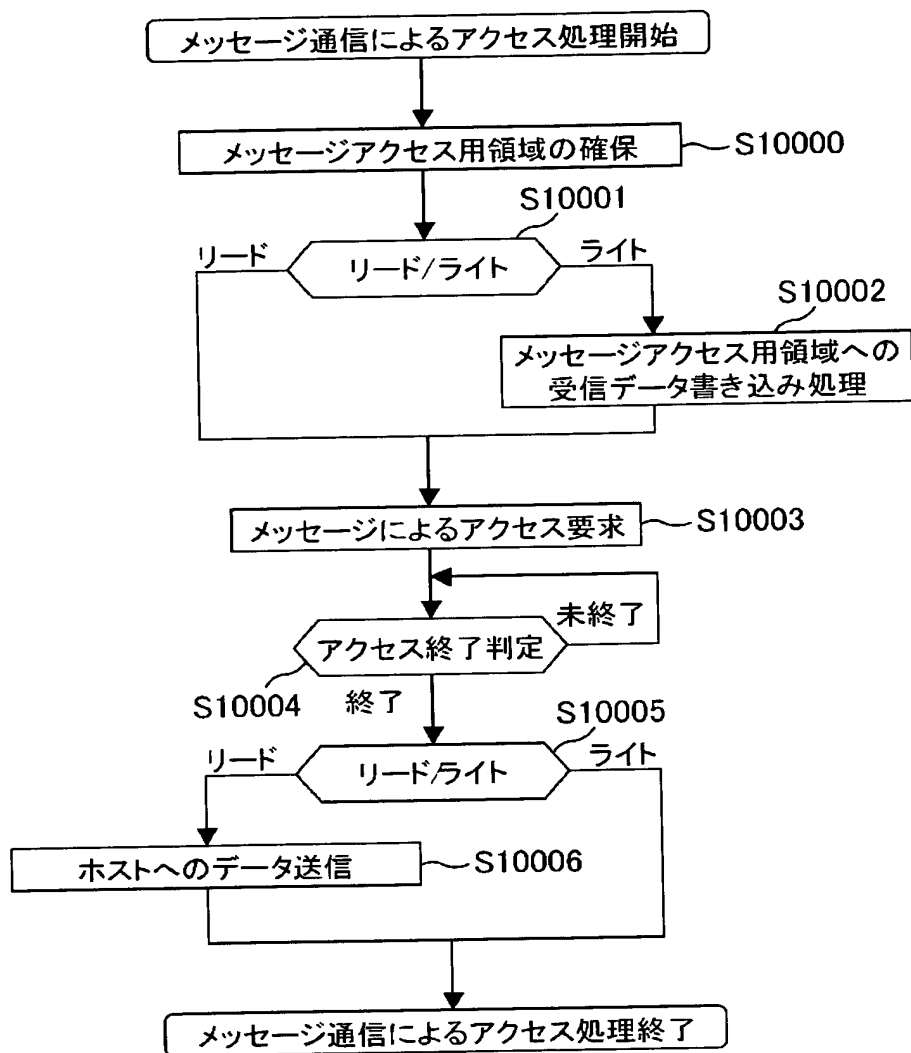
【図 3 3】



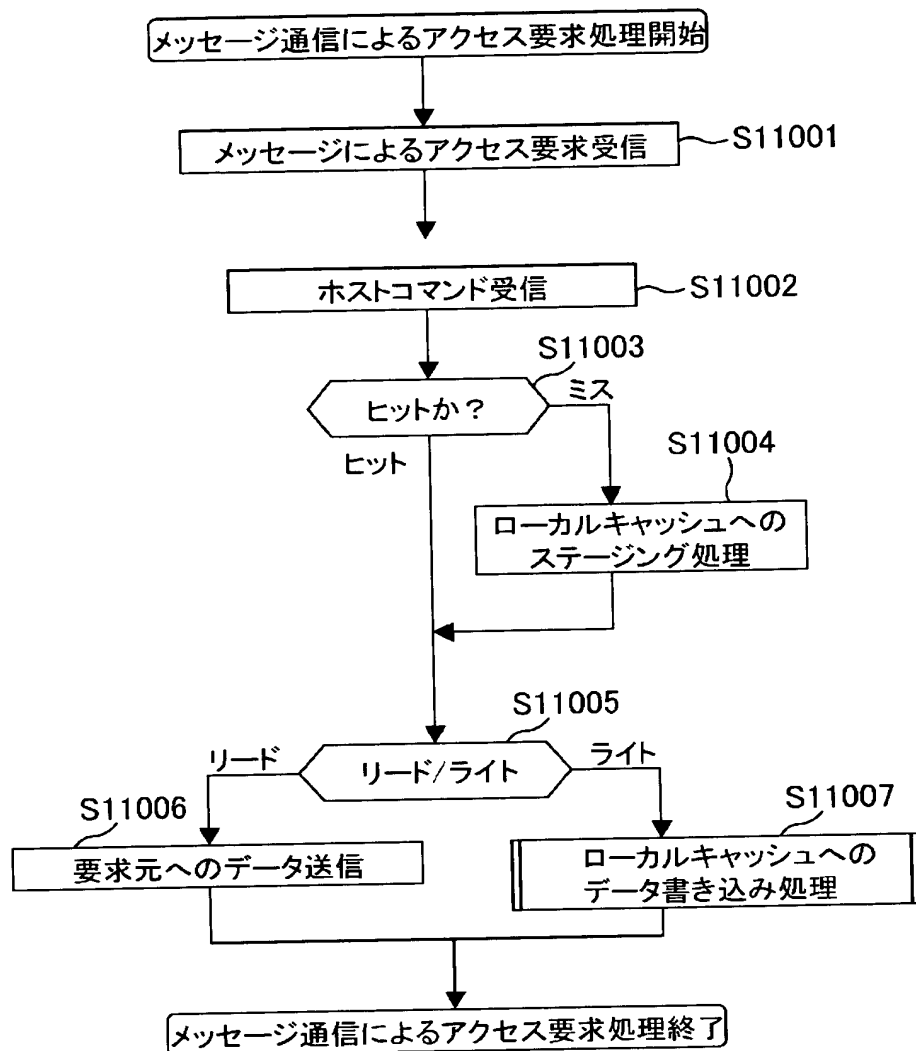
【図 3 4】



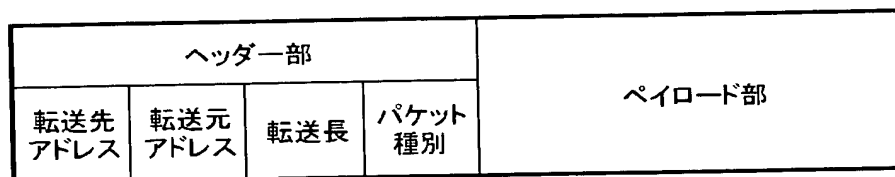
【図 35】



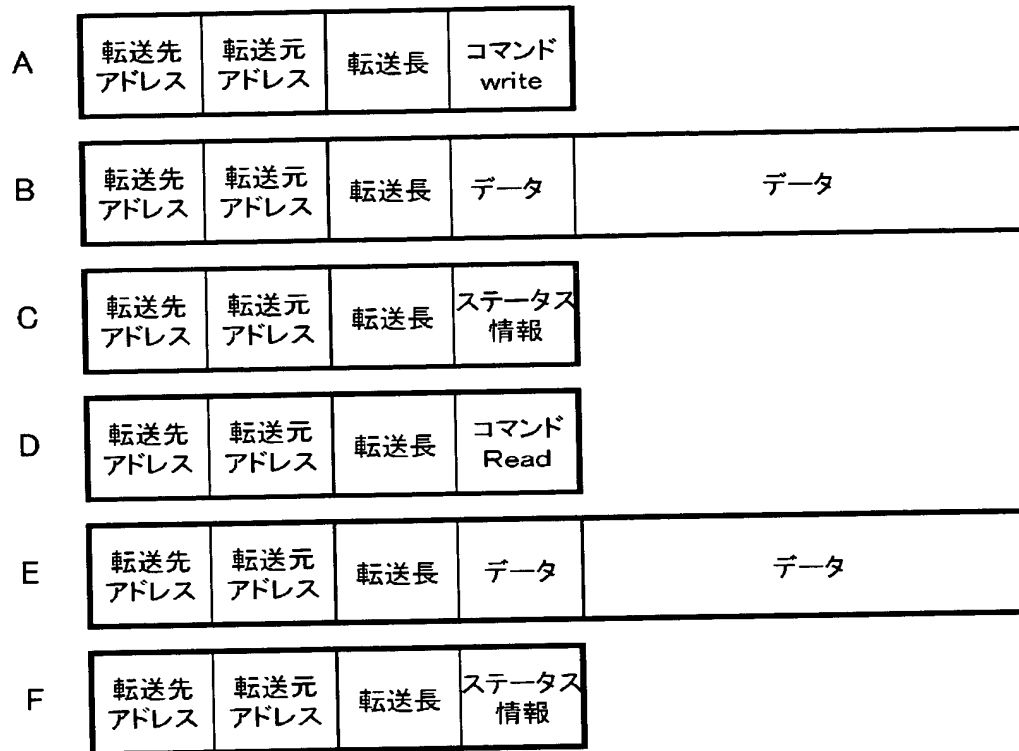
【図 3 6】



【図 3 7】



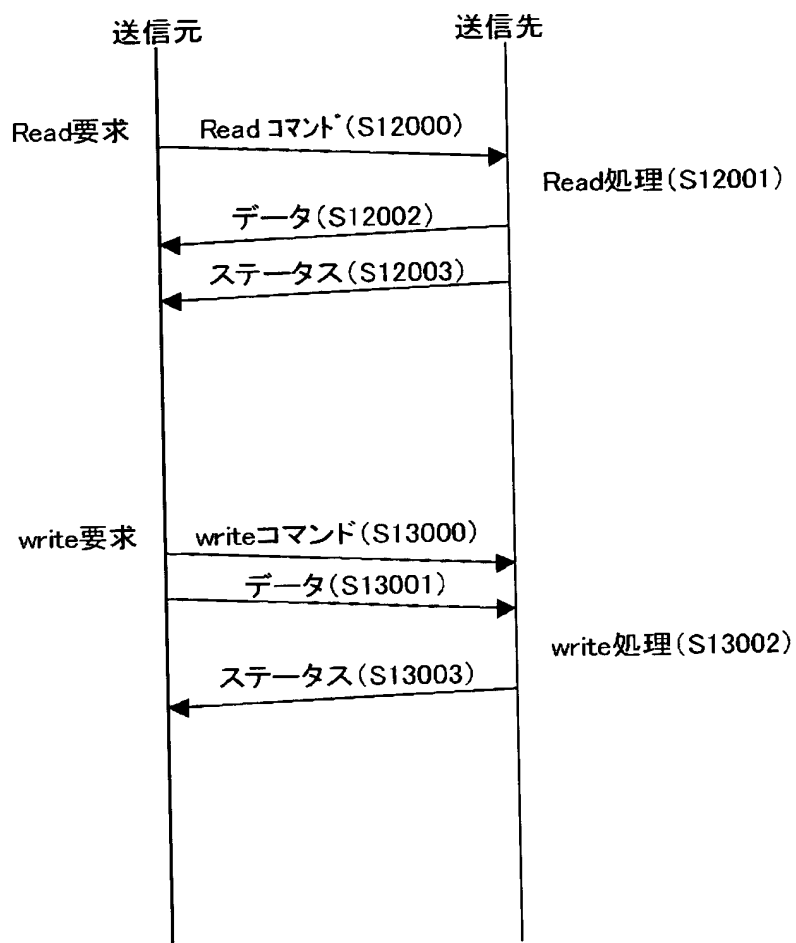
【図 38】



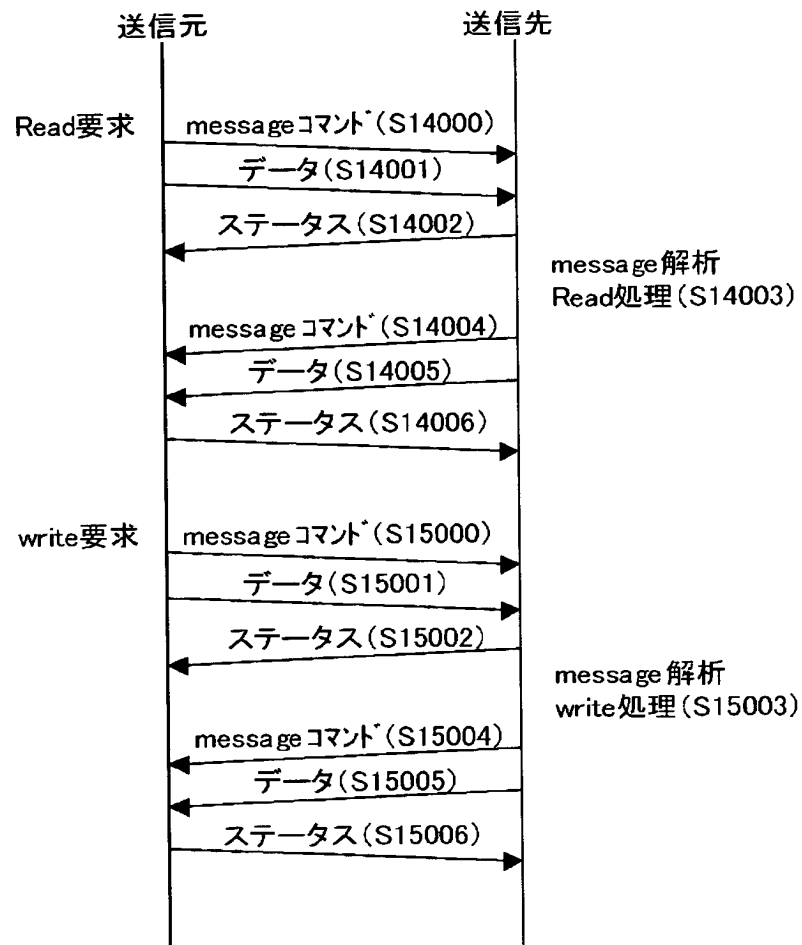
【図 39】



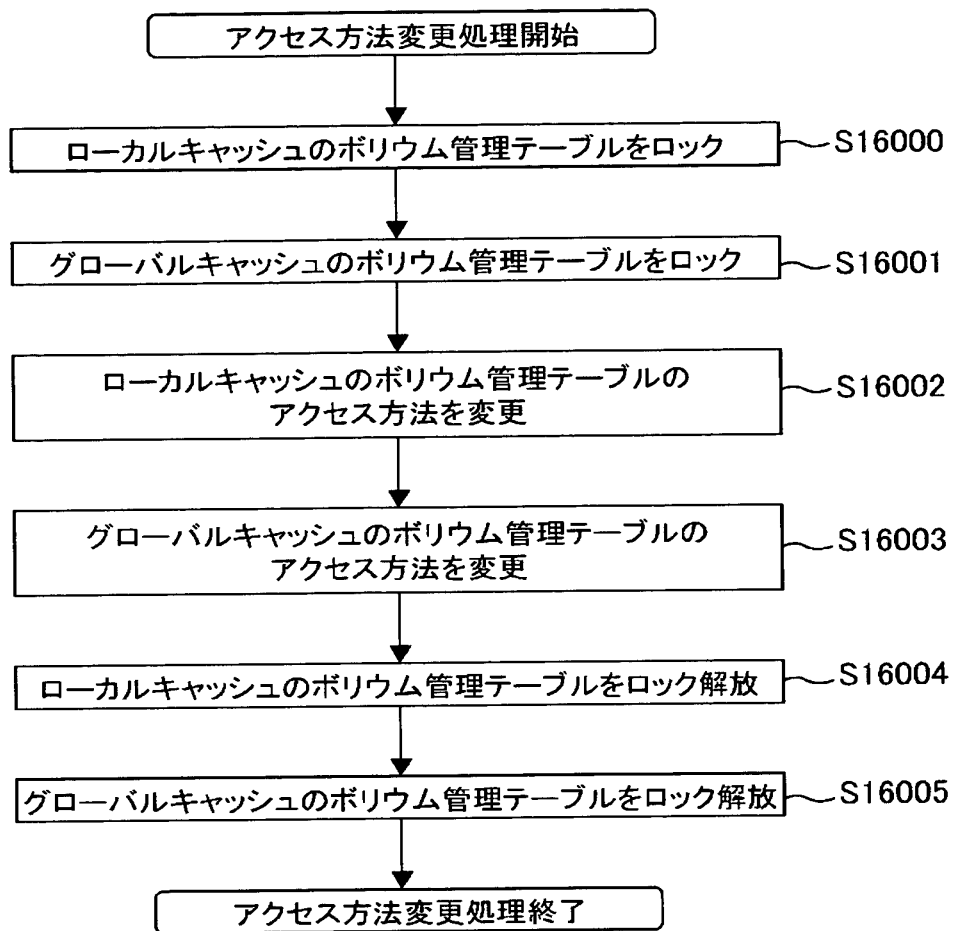
【図 40】



【図 4 1】



【図 4 2】



【書類名】 要約書

【要約】

【解決手段】 データ入出力要求を受信するためのホストインタフェース制御部が形成されたチャンネル制御ユニットと、前記データ入出力要求に応じて、データを記憶するための記憶ボリュームに対する前記データの入出力制御を行うためのディスクインタフェース制御部が形成されたディスク制御ユニットと、前記データを記憶するためのメモリが形成されたキャッシュメモリユニットと、前記ホストインタフェース制御部と前記ディスクインタフェース制御部と前記メモリとが形成されたストレージ制御ユニットとを挿抜可能な装着部と、前記チャンネル制御ユニット、前記ディスク制御ユニット、前記キャッシュメモリユニット、及び前記ストレージ制御ユニットを通信可能に接続する内部接続部とを備えることを特徴とする記憶デバイス制御装置に関する。

【選択図】 図 2

特願 2 0 0 3 - 1 1 1 4 0 5

出 願 人 履 歴 情 報

識別番号

[0 0 0 0 0 5 1 0 8]

1 . 変更年月日

1 9 9 0 年 8 月 3 1 日

[変更理由]

新規登録

住 所

東京都千代田区神田駿河台 4 丁目 6 番地

氏 名

株式会社日立製作所